



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA



Tesis de Máster

Máster de Posgrado Oficial en Ingeniería del Software, Métodos
Formales y Sistemas de Información

Modelado e Implementación de las Probabilidades de Riesgo al Modelo Conceptual del Genoma Humano

Julio de 2013, Valencia

Alexander Santana Donato
alsan6@posgrado.upv.es

Director: Oscar Pastor López
opastor@pros.upv.es

Co-directora: Mercedes R. Fernández Alcalá
mfernandez@pros.upv.es

Índice General

Agradecimientos.....	4
1. Introducción.....	5
1.1 Motivación.....	5
1.2 Estructura de la tesis.....	7
2. Planteamiento del problema.....	9
2.1 Selección de enfermedad.....	10
2.2 Objetivos.....	18
3. Estado del arte.....	19
4. Solución del problema.....	29
4.1 Especificación de genes y variaciones relacionadas con la AGA.....	29
4.2 Desarrollo del proceso de probabilidades de riesgo.....	33
4.3 Ampliación del modelo conceptual.....	37
4.3.1 Modelo de las probabilidades de riesgo.....	37
4.3.2 Soporte de sinónimos de RS en el modelo conceptual.....	39
4.4 Carga de variaciones.....	40
4.5 Explotación de contenidos.....	59
5. Conclusión.....	61
6. Referencias bibliográficas.....	62
7. Anexos.....	65

Agradecimientos

Primeramente quiero agradecer, de manera muy especial, a Jehová Dios porque Él ha hecho posible la realización de mis estudios aquí en España, y ha estado a mi lado todo este tiempo. Gracias papito Dios.

También quiero agradecer al Centro de Investigación en Métodos de Producción de Software (Pros) y a todos los miembros del proyecto genoma por todos los momentos chéveres que compartimos y su apoyo en todo este trayecto.

Quiero hacer una mención especial a mi director de tesis, Oscar Pastor, por haberme dado la oportunidad de trabajar en este proyecto y por su ayuda en el desarrollo de esta tesis.

Agradezco a mi co-directora Mercedes R. Fernández por haberle dado el seguimiento a mi tesis y por haber estado disponible en todo tiempo. También agradezco a Ana Levin por su disponibilidad en el área biológica. Estoy muy agradecido de ambas por haber creado un buen ambiente de trabajo entre todos mis compañeros.

Hago una mención muy especial a mi novia y prometida, la cual amo con todo mi corazón, porque a pesar de los miles de kilómetros que nos separan, ha estado presente y me ha apoyado todo este tiempo.

Estoy muy agradecido de mis padres por haber confiado en mí, y aun estando a la distancia, me han brindado su apoyo animándome cada día a seguir luchando en mis estudios. Son especiales para mí.

Les doy las gracias a mis compañeros de máster con los que compartí todo este recorrido, y de manera especial, a mis compañeros y amigos (como hermanos) Fausto Nelson y José Reyes con los cuales pasé grandes momentos inolvidables y con los que he hecho grandes lazos desde mi llegada a España; con su apoyo y compañía han hecho que este tiempo sea de mucho provecho. Sois una bendición para mí vida.

Quiero hacer una mención a mi amiga y hermanita Diana Vanesa Buitrago porque su cariño, su compañía y sus palabras me han servido de motivación en todo este tiempo.

Por último les doy las gracias a mis amigos y a mis pastores de la iglesia de Dios en Cristo de Republica Dominicana, y a mis amigos y a mis pastores de la iglesia Rey de Reyes de Valencia porque siempre han creído en mí y me han ofrecido su apoyo durante todo el desarrollo de este máster y de esta tesis.

1. Introducción

1.1 Motivación

Existen muchas enfermedades que se encuentran escritas en los genes, algunas se manifiestan desde el principio de la vida de la persona y otras empiezan a causar sus efectos a lo largo de los años. Muchas de estas enfermedades genéticas, si se detectan a tiempo, pueden curarse y evitar daños graves.

Las enfermedades genéticas presentan rasgos fenotípicos, que pueden ser físicos o conductuales, asociados al genotipo. El genotipo es la información genética, encontrada en el ADN, que se manifiesta físicamente en cada individuo determinando el color de los ojos, el color de la piel, el tipo y color del pelo, entre otras; estas manifestaciones del genotipo se conocen como fenotipo.

Muchas veces en el ADN ocurren ciertas alteraciones conocidas como mutación. Las mutaciones por lo general producen enfermedades, leves o graves, en un individuo como por ejemplo: un cáncer, una malformación, el síndrome de Down, etc., pero en otras ocasiones los cambios no son perjudiciales, ya que hacen que un individuo sea diferente de otro como por ejemplo: el color de los ojos, la estatura, el tipo de piel, etc.

Debido que muchas enfermedades genéticas son tan comunes, con el tiempo se han hecho numerosos estudios en los genes para analizar más a fondo las mutaciones y así poder determinar que partes del código genético están relacionadas con ciertos rasgos fenotípicos de interés, y que tan fuerte es esa relación [7] [11] [16]...

Mientras más se avanza en el estudio de las mutaciones y sus efectos, se va obteniendo un mayor volumen de datos biológicos. El volumen de datos es tan grande que fue necesaria la aparición de potentes bases de datos que pudieran contener y administrar toda esta información.

Una vez empezó a haber suficiente información sobre las mutaciones y sus efectos, surgió la aparición de empresas que pueden ofrecer pruebas (test) genéticas, esto lo hacen analizando el código genético de una persona para determinar si en dicho código se encuentran las mutaciones asociadas a una determinada enfermedad. Con el test genético las personas tienen la posibilidad de tratar con tiempo la enfermedad genética a la cual esta predispuesto y que pudiera conducirles a la muerte o a condiciones catastróficas.

Entre tantas enfermedades genéticas está la Alopecia Androgénica que es una enfermedad genética que consiste en la caída del pelo de manera progresiva hasta dejar el cuero

cabelludo, en muchas ocasiones, totalmente al descubierto. Esta enfermedad se presenta en dos patrones que son: alopecia de patrón masculino y alopecia de patrón femenino las cuales afectan a las personas independientemente del sexo. La padecen muchos seres humanos, y los hombres son más afectados que las mujeres.



Nos centraremos en esta enfermedad ya que es muy común en la sociedad, especificaremos las distintas variaciones¹ relacionadas con esta, y determinaremos la probabilidad que tiene una persona en padecerla. Para lograr esto, utilizaremos la metodología SILE desarrollada por el centro de investigación en métodos de producción de software Pros de la UPV.

¹ En esta tesis utilizaremos este nombre para referirnos a las mutaciones.

1.2 Estructura de la tesis

Esta tesis se desarrollamos con la siguiente estructura

Planteamiento del problema

En este capítulo se presenta el problema existente que se quiere resolver en el desarrollo de esta tesis.

- Veremos que la alopecia androgenética necesita estar bien identificada obteniendo toda la información valiosa que se encuentra dispersa en las distintas bases de datos y en distintos papers.
- También es mostraremos la importancia de desarrollar un proceso probabilístico que nos permita determinar la probabilidad que tienen las variaciones para que se manifiesten fenotípicamente. En nuestro caso tomaremos las variaciones que están relacionadas con la predisposición al padecimiento de la Alopecia Androgenética.

Presentamos tres enfermedades que son: la enfermedad de Alzheimer, Síndrome de Down y la Alopecia Androgenética. Especificaremos el motivo por el cual hemos seleccionado la Alopecia Androgenética y mostraremos los objetivos perseguidos que nos llevan a solucionar el problema.

Estado del arte

Presentaremos en qué consiste la metodología SILE y hablaremos de distintas bases de datos genómicas que hemos consultado como son: NCBI, Genecard, Ensembl. También hablaremos de algunas empresas que ofertan test genético como son: 23andme, IB Biotech y TRK Genetic.

Además presentaremos una wiki que ofrece información muy valiosa sobre los SNPs llamada SNPedia.

Solución del problema

Este capítulo de desarrolla de la siguiente manera

- Se muestra una lista con los genes y Variaciones encontrados que están relacionados con la predisposición al padecimiento de la Alopecia Androgenética.

Además validamos estas variaciones para identificar solo las variaciones que son más relevantes en el padecimiento de la Alopecia androgenética.

- Desarrollamos el proceso que se lleva a cabo para determinar las probabilidades de las variaciones.

Una vez desarrollado este proceso, especificaremos las probabilidades que tiene cada una de las variaciones identificadas que se relacionan con la Alopecia Androgenética.

- Actualizaremos el modelo conceptual, desarrollado por el Centro de Investigación en Métodos de Producción de Software (PROS), ampliándolo y modelando el proceso probabilístico que determina el peso de una variación.
- Para poder cargar los genes y variaciones fue necesario ampliar la base de datos adecuándola a la nueva actualización del modelo conceptual.

Mostraremos cada una de las tablas añadidas y ampliadas, su estructura, el esquema de base de datos y las distintas restricciones para que los datos se mantengan con integridad.

Conclusión

En este capítulo concluimos habiendo integrado el modelo del proceso probabilístico para las variaciones genéticas.

Para lograr esta integración, identificamos los genes y variaciones más importantes que se relacionan con la Alopecia Androgenética, a partir de las cuales mostramos todos los pasos a dar para obtener las probabilidades de cada una de las variaciones.

Mostramos algunos trabajos futuros que permitirán seguir profundizando y avanzando en la relación genotipo-fenotipo y sus probabilidades.

2. Planteamiento del problema

Debido al avance en la genética las personas cada vez están más interesadas en conocer su código genético para poder detectar a tiempo cualquier enfermedad genética que esté escrita en sus genes y así poder tratarla y evitar que se le manifieste o al menos reducir los daños que les pueda traer la enfermedad.

Para que las empresas activas en el campo del testado genético puedan ofrecer información fiable sobre la predisposición al padecimiento de una enfermedad genética determinada, sus conclusiones necesitan estar fundamentadas en estudios que profundizan en todas las variaciones que están relacionadas a esa enfermedad.

Esto es debido a que una enfermedad genética puede estar asociada a una o muchas variaciones.

Las enfermedades genéticas son manifestadas por las variaciones de la siguiente manera:

- Existen enfermedades que se manifiestan solo cuando se da una combinación de variaciones en específico llamadas Haplotipos². Esto quiere decir que si aparece una parte de esta combinación, la enfermedad no se manifiesta.

Un ejemplo de enfermedades que utilizan haplotipos es la Sensibilidad al Alcohol. El **gabra2** es un gen codificante a proteína y que está relacionado con el alcoholismo [21]; contiene tres SNPs³ que conforman un haplotipo rs279871, rs279836 y rs279845.

- Otras enfermedades están relacionadas con un conjunto x de variaciones, pero para manifestarse no es necesario que estén todas estas variaciones en el código genético de la persona, con el hecho de que haya una sola variación del conjunto x la enfermedad podría manifestarse o no; incluso con el conjunto x completo la enfermedad puede no manifestarse. Esto quiere decir que estas variaciones no aseguran en la totalidad la manifestación de la enfermedad sino que están determinada por una probabilidad de riesgo.

² Es un conjunto de SNPs que por estar cerca uno del otro, están asociados y en **general** se heredan en conjunto.

³ Single Nucleotide Polymorphism.

- Algunas enfermedades tienen una relación tan fuerte con una variación que siempre que dicha variación aparece en el código genético, se manifiesta la enfermedad.

Una de estas enfermedades es la intolerancia a la lactosa, la cual está relacionada con el gen LCT que codifica la proteína lactasa.

De manera que para hacer un test genético fiable es necesario tener estos puntos muy claros. Uno de los puntos más importantes y difíciles de obtener es la determinación de las **probabilidades** de riesgo que ofrece cada variación.

La dificultad de de las probabilidades de riesgo radica en:

- Mientras más variaciones se encuentran en una persona determinada, existe un mayor riesgo a padecer la enfermedad relacionada a esas variaciones. De manera que cada variación aporta un aumento en el nivel de probabilidad de riesgo al padecimiento de dicha enfermedad.
- Es necesario saber la probabilidad de cada variación para saber cuáles son las de mayor riesgo y cuáles las de menor riesgo.
- La probabilidad de riesgo de una persona depende de la población a la que pertenezca, ya que una variación puede manifestarse fenotípicamente en una población más que en otra.

Las personas interesadas en hacerse un test genético se sienten más seguras cuando la información que obtienen es más directa, reciben con más agrado y conformidad que se les dé un porcentaje x de padecer una enfermedad determinada, a que se les diga tan solo que está predispuesto a padecerla.

2.1 Selección de enfermedad

Debido a la importancia de las probabilidades de riesgo anteriormente mencionadas, nos interesa desarrollar todos los pasos a seguir para poder determinarlas, así como modelar todo este proceso en un modelo conceptual el cual especificaremos más adelante.

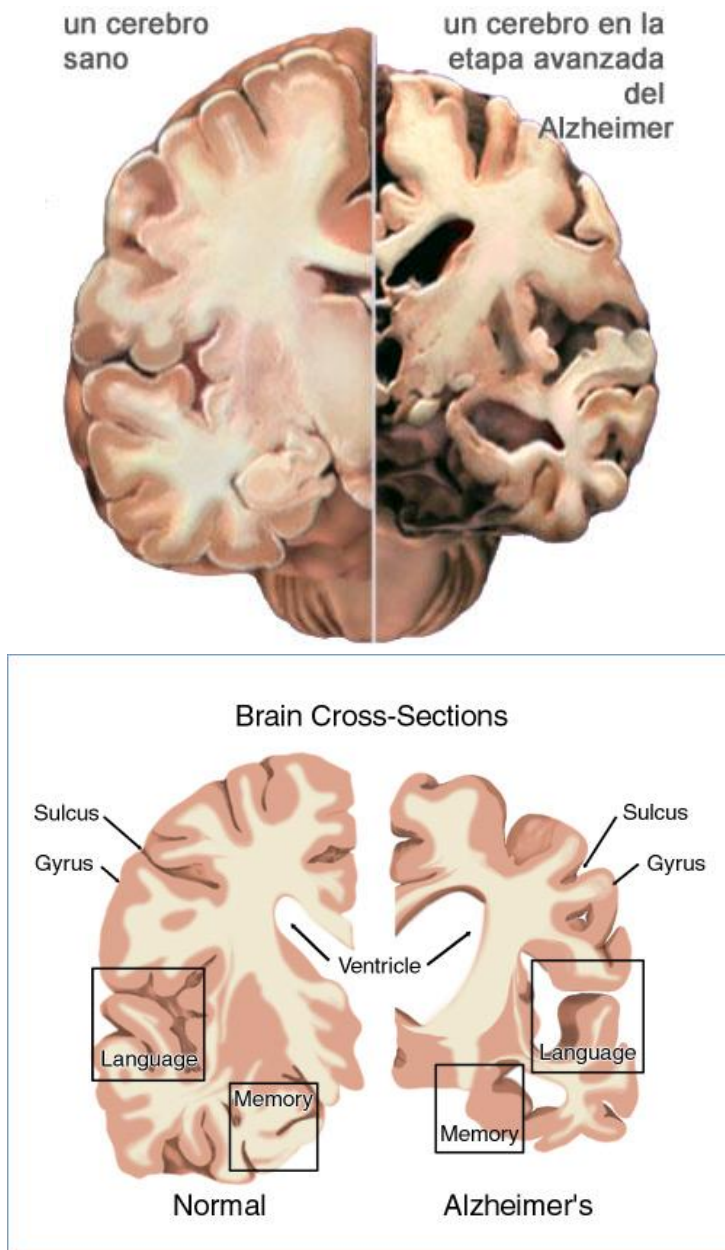
Estos pasos son aplicables a muchas enfermedades, pero seleccionaremos una para poder desarrollar todo este proceso y así poder mostrar la aplicabilidad del proceso probabilístico en las enfermedades genéticas.

Analizamos tres enfermedades genéticas de entre las cuales hemos elegido una para integrar toda su información genética, así como determinar las condiciones que se deben cumplir para que una persona tenga mayor o menor probabilidad de riesgo a padecerla. A continuación describiremos las enfermedades estudiadas para la selección.

1. Enfermedad de Alzheimer

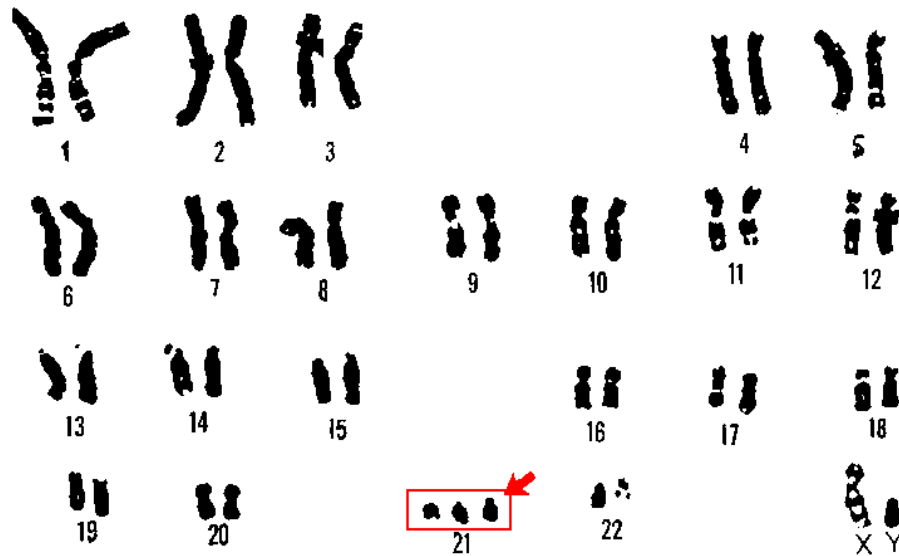
Es un tipo de demencia, la mayoría de los casos de demencia son ocasionados por el Alzheimer. Las neuronas mueren y el cerebro se deteriora causando la pérdida de la memoria y el intelecto. Por el momento esta enfermedad no tienen cura [17], pero existen terapias con el objetivo de que el deterioro del cerebro sea mucho más lento [18].

Las personas más afectadas son aquellas que tienen más de 65 años. Fue identificada por Emil Kraepelin y Alois Alzheimer en 1906 tras haberle dado seguimiento durante 5 años a una mujer con esta enfermedad. Cuando dicha mujer murió Alois Alzheimer examinó su cerebro y fue cuando esta enfermedad fue totalmente identificada, y Kraepelin decidió bautizarla con el nombre **Alzheimer**.



2. Síndrome de Down

También es conocido como trisomía 21 [19], esta ocurre cuando aparece una copia adicional del cromosoma 21. Las personas afectadas con este síndrome tienen características físicas similares que son reconocibles, además presentan retrasos mentales así como anomalías en el desarrollo físico. Los efectos de esta enfermedad pueden ser muy graves o muy leves.



3. Alopecia androgenética (AGA)

La alopecia es la pérdida total o parcial del pelo, dejando todo o parte del cuero cabelludo al descubierto. La alopecia, que es lo mismo que calvicie, puede ocurrir en cualquier lugar del cuerpo, y es mucho más común en hombre que en mujeres.

La mayoría de los casos de pérdida de pelo están determinados por problemas genéticos. La alopecia más común es la **Alopecia Androgénica** (también llamada Alopecia androgénica, calvicie común o AGA) la cual afecta a un 95% de las personas con problemas de calvicie. Una persona puede ser afectada por la calvicie común en cualquier etapa de la vida. Normalmente empieza entre los 18 y 30 años, y afecta a más del 90% de las personas a partir de los 75 años, de modo que, el riesgo aumenta con la edad.

Tanto el estrés como la mala alimentación aumentan el riesgo de sufrir esta enfermedad.

Tratamientos

Hay que recordar que esta enfermedad es genética y se necesitan tratamientos especiales que pueden dar muy buenos resultados si se aplican en fases tempranas o medias de la AGA⁴.

Se necesita estudiar cada caso específicamente para poder determinar cuál sería el tratamiento que se adecue a la etapa de calvicie en que se encuentre la persona.

Mientras más temprano se apliquen los tratamientos, se podrán obtener mejores resultados

Patrones de la Alopecia Androgenética (AGA)

Existen dos patrones de AGA (AGA de patrón masculino y AGA de patrón femenino) los cuales se diferencian uno de otro por la forma en que aparecen y se extienden. El AGA de patrón masculino es mucho más común que el de patrón femenino. Los patrones actúan en las personas independientemente del sexo, de manera que, un hombre puede estar afectado por AGA de patrón femenino o masculino y una mujer puede ser afectada por AGA de patrón masculino o femenino.

AGA de Patrón Masculino

Inicia con una pérdida mínima de pelo en la parte frontal del cuero cabelludo, que va avanzando para formar las entradas; con el pasar del tiempo, la parte de la coronilla empieza a quedarse sin cabello y se extiende hasta unirse con las entradas dejando una sola calva, la cual sigue expandiéndose hacia los laterales y hacia la

⁴ Nombre atribuido a la Alopecia Androgenética

parte posterior; Normalmente, en la etapa final solo queda una delgada franja que se extiende de una oreja a otra, y en ocasiones el cuero cabelludo se queda totalmente descubierto de pelo.

Escala de Hamilton - Norwood

La escala de **Hamilton-Norwood**, también conocida como escala de Hamilton o escala de Norwood o Escala de NW, fue creada por James Hamilton en los años 50s y actualizada o mejorada por O'Tar Norwood en los años 70s. Consiste en siete estadios por los cuales pasa una persona con alopecia androgenética (AGA) con patrón masculino, permitiendo tener un mejor control del desarrollo de esta enfermedad para poder tomar decisiones sobre el tipo de tratamiento que se puede aplicar para contrarrestar los efectos y el avance del AGA y en casos avanzados, optar por las intervenciones quirúrgicas. Ni las intervenciones quirúrgicas, ni ningún tratamiento en la actualidad obtienen buenos resultados cuando la AGA se encuentra en el estadio VII.

1. **Tipo I:** La pérdida del pelo empieza en la zona frontal; en esta etapa, la caída del cabello es difícil de apreciar ya que es mínima. Normalmente en este estado, los tratamientos hacen un buen trabajo.



2. **Tipo II:** El pelo de la zona frontal y temporal se debilita y se cae dejando las famosas entradas. En esta etapa es cuando se empieza a notar con claridad la pérdida del cabello. En este estado, los tratamientos son muy efectivos.



3. **Tipo III:** Las entradas se acentúan más y empieza a desaparecer el pelo de la zona de la coronilla, ampliándose con el paso del tiempo. Esta es la primera etapa que es considerada como calvicie o alopecia. Llegados a este punto, los tratamientos siguen dando muy buenos resultados.



4. **Tipo IV:** La coronilla se amplía mucho más que en el tipo III, así como la zona frontal y temporal. En la parte superior de la cabeza se encuentra muy poco cabello y en ocasiones no se encuentra ningún pelo. En esta etapa es necesario

aplicar tratamiento de manera urgente, los cuales aún pueden dar buenos resultados.



5. **Tipo V:** La franja que divide la coronilla de las entradas empieza a desaparecer uniendo, poco a poco, estas dos zonas del cuero cabelludo. Las intervenciones quirúrgicas son necesarias, ya que los tratamientos son inútiles en esta etapa.



6. **Tipos VI:** La coronilla se une totalmente con la parte de las entradas, creándose una sola calva y siguiendo su curso hacia los laterales.



7. **Tipo VII:** Es la etapa más avanzada de la calvicie en la cual solo queda una pequeña franja en la parte posterior llegando hasta un poco por encima de las dos orejas. Llegados a este punto, no existe solución posible.



Los hombres son los que se ven más afectados por la alopecia de patrón masculino, pero también existe un número muy pequeño de mujeres afectadas con este patrón.

AGA de Patrón Femenino

El pelo comienza a debilitarse en la zona occipital, la coronilla y laterales; pero no necesariamente estas tres zonas estarán afectadas. En la mayoría de los casos la zona afectada es la de la coronilla "Parietal". Generalmente no provoca la pérdida completa o casi completa del pelo como ocurre en el patrón masculino.

Escala de Ludwig

En los años 70s el Dr. Ludwig creó una escala para denotar el avance de la alopecia de patrón femenino. Ludwig identificó tres etapas o tipos.

1. **Tipo I:** Esta considerada como una caída suave del pelo, ya que está dispersa y es muy poco notoria.



2. **Tipo II:** El cabello se debilita más y se hacen mucho más finos. En este punto se puede notar la falta de cabello con más facilidad que en la etapa anterior, la cual empieza a extenderse en la parte de la coronilla. Esta etapa se considera como una caída moderada de pelo.



3. **Tipo III:** Es la última etapa de este tipo de calvicie y se considera como extensa. En esta etapa la calvicie se extiende por toda la coronilla dejándola totalmente expuesta y notoria.



La AGA de patrón femenino es mucho menos común que el patrón masculino y generalmente se encuentra en las mujeres, pero también existen unos pocos casos de hombre que la contienen.

Existe mucha importancia en que la AGA sea detectada a tiempo pues de esta manera los tratamientos podrán brindar un mejor resultado, la alopecia androgenética puede llegar a un estado en el cual no hay tratamiento que pueda ser útil para contrarrestar sus efectos.

Entre las anteriormente mencionadas hemos seleccionado la Alopecia Androgenética por varias razones:

1. Cuenta con alrededor de 25 variaciones que se adecuan muy bien a las probabilidades de riesgo anteriormente mencionadas. Más adelante estaremos comentando sobre estas variaciones.
2. Puesto que es una enfermedad muy común en la sociedad. Aunque la raza blanca es la más afectada, las demás razas también tienen riesgo de sufrir esta enfermedad. De manera que existen una gran cantidad de personas que son afectadas por la Alopecia Androgenética, y mientras más envejecen las personas, mayor riesgo tienen a sufrirla.
3. A diferencia de la Enfermedad de Alzheimer y del Síndrome de Down, la AGA si tiene cura.

Cuando la AGA es detectada y diagnosticada en fases tempranas o medias de su desarrollo, los tratamientos pueden dar unos resultados sorprendentes.

4. Muchos hombres jóvenes la padecen. Los hombres de raza blanca de 30 y 50 años son afectados por esta enfermedad en un 25 y 50% respectivamente.
5. Muchas personas con Alopecia Androgénica⁵, especialmente las mujeres y hombres jóvenes, tienen problemas de baja autoestima.

Por estas razones creemos que la AGA nos sería muy útil para crear un servicio web de test genético haciendo uso de las probabilidades de riesgo.

⁵ Es lo mismo que Alopecia Androgenética.

2.2 Objetivo

El objetivo general de esta tesis es desarrollar y modelar el proceso de las probabilidades de riesgos de las variaciones genéticas de cualquier enfermedad genética.

Recordar que utilizaremos la Alopecia Androgenética para desarrollar las probabilidades de riesgo de sus variaciones, sabiendo que las probabilidades de riesgo son aplicables a cualquier enfermedad genética que lo requiera.

Para poder cumplir con el objetivo general, hemos desarrollado los siguientes objetivos específicos.

- Obtener todos los genes que se relacionan con la Alopecia Androgenética e identificar los que son más relevantes ya que están validados por estudios que se han hecho en distintas poblaciones.
- Obtener los RSs⁶ o variaciones de los genes ya identificados anteriormente, que determinan la predisposición al padecimiento de la Alopecia Androgenética.
- Identificar la probabilidad de riesgo de cada RS relacionado con la Alopecia Androgenética.
- Modelar todo el proceso de probabilidades de riesgo a partir del modelo conceptual, creado por el Centro de Investigación en Métodos de Producción de Software (Pros) de la UPV, para que éste pueda contemplar las probabilidades de riesgo de cualquier enfermedad genética.
- Actualizar la definición de la base de datos para que se adapte a los nuevos cambios del modelo conceptual; de esta manera se podrá administrar y gestionar las variaciones y sus probabilidades de riesgo.

Para poder cumplir estos objetivos seguiremos los pasos de la metodología SILE para así poder obtener toda la información que necesitamos; mas adelante estaremos definiendo esta metodología.

Una vez cumplidos todos estos objetivos podremos ofrecer información más fiable sobre la predisposición al padecimiento de cualquier enfermedad genética, en este caso, de la Alopecia Androgenética que puede tener un individuo, ya que observaremos una cantidad considerable de variaciones relacionadas con la AGA y obtendremos así sus probabilidades de riesgo para el padecimiento de ésta.

⁶ RS es un número de identificación único que utiliza NCBI para referirse a las variaciones o mutaciones. En esta tesis trataremos el RS como Variación y viceversa.

3. Estado del arte

Existen muchas bases de datos cargada de información genética, llamadas bases de datos genómicas, y en ocasiones estas bases de datos tienen rasgos fenotípicos asociados a su genotipo, pero no siempre es así.

Existen también empresas que ofrecen tests genéticos para distintas enfermedades genéticas sobre las que ellos han estudiado.

23andme

Fue fundada en 2006 por Linda Avey, Paul Cusenza y Anne Wojcicki. Es una empresa que realiza diagnósticos genéticos utilizando un kit de testado genómico para determinar las variaciones genéticas de un individuo con respecto a una enfermedad determinada.

23andme también brinda información de cada una de sus variaciones y permite hacer diferentes tipos de búsquedas.

Al hacer una búsqueda por genes nos muestra una lista de variaciones asociadas a dicho gen. También permite la búsqueda por cromosoma desplegando una lista donde muestra todos los genes y variaciones asociadas a dicho cromosoma.

Un **SNP**⁷, también llamado polimorfismo⁸ de un solo nucleótido, es una variación que afecta a un solo nucleótido. Cuando las alteraciones son de muy pocos nucleótidos de la secuencia, también se consideran SNP.

Encontramos la posibilidad de buscar por SNP. Introduciendo la variación que deseamos en el campo SNP, podemos observar el gen al que pertenece la variación, los alelos⁹ y algunos links de referencia.

- **Entrez Gene:** Nos dirige a NCBI para darnos más detalles sobre el gen al que pertenece la variación que nos interesa.
- **Google Scholar (gene):** Nos lleva a Google Scholar para mostrarnos una serie de artículos relacionados con el gen en cuestión.

⁷ Single Nucleotide Polymorphism.

⁸ Cuando en una población existen muchos alelos de un mismo gen.

⁹ Es la variante de un gen

- **dbSNP Lookup:** Al hacer click aquí nos dirige a NCBI donde obtenemos información más amplia y detallada sobre el SNP que estamos buscando.
- **Google Scholar (SNP):** Abre el Google Scholar con una serie de papers relacionados con el SNP en cuestión.

En la siguiente imagen se muestra como 23andme hace las búsquedas.

Showing raw data for SNP **rs2003046**, which is on chromosome **1**.

Jump to a gene **Go** a SNP: **Go**

or a chromosome:

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 X Y MT

« Return to your whole genome.

Gene	Position	SNP	Versions	Genotypes
C1orf127	11032827	rs2003046	A or C	CC
				Lilly Mendel (Mom)
				Greg Mendel (Dad)

Reference Links:

Entrez Gene

Google Scholar (Gene)

dbSNP Lookup

Google Scholar (SNP)

Al analizar 23andme con respecto a la alopecia androgénica, hemos observado que utilizan cuatro variaciones [9].

En la siguiente tabla mostramos las variaciones, los genes y los nombres de los genes que utiliza 23andme.

Lista de Variaciones de 23andme

#	Nombre	Gen	RS
1	Chromosome 1 Open Reading frame 127	C1orf127 cerca del gen TARDBP	rs2003046
2	Histone Deacetylase 4	Intergenic cerca del gen HDAC4	rs9287638
3	Histone Deacetylase 9	HDAC9	rs2073963
4	Solute Carrier Family 14 (Urea Transporter), Member 2	SLC14A2 cerca del gen SETBP1	rs8085664

23andme muestra la población a la que se le hizo el estudio, la cantidad de personas enfermas, la cantidad de personas sanas y el Odds Ratio¹⁰ de cada una de las cuatro variaciones que ellos utilizan.

¹⁰ También llamado OR, Es una medida estadística.

IB Biotech

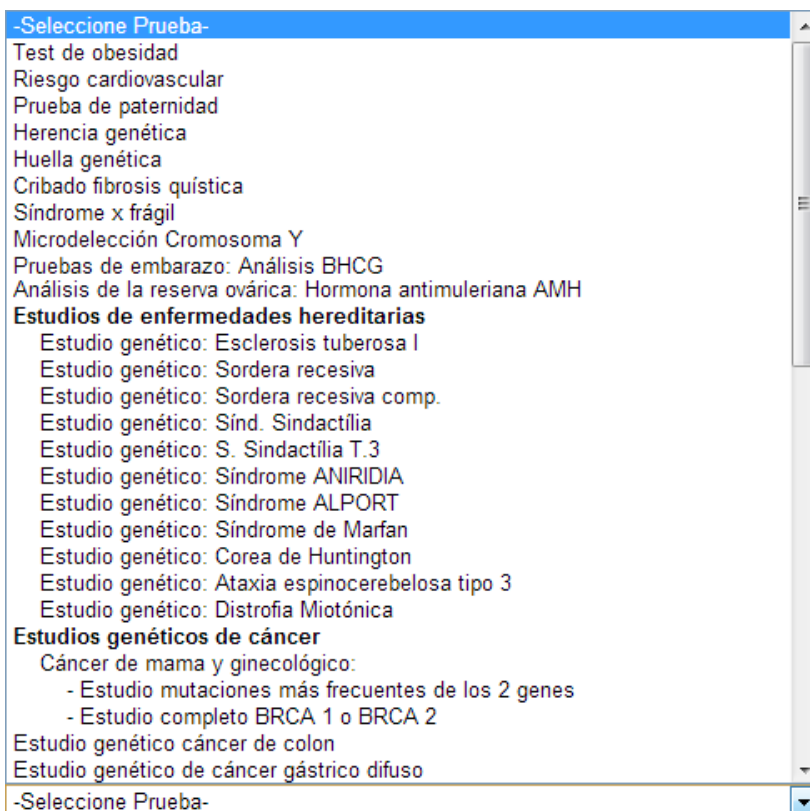
Es una empresa activa en el campo del test genético con el fin de ofrecer medidas preventivas que permitan contrarrestar el desarrollo de algunas enfermedades genéticas.

Cuenta con una amplia lista de servicio de pruebas genéticas que realizan utilizando un kit para tomar la muestra genética. Además detallan paso el modo de empleo del kit.

En la siguiente imagen mostramos una pequeña vista de IB Biotech. Al hacer click en “Genética desde casa, tú mismo” nos permite iniciar el proceso de solicitud de prueba y kit, y nos permite seleccionar el tipo de prueba de una amplia lista.



Lista de pruebas



TRK Genetics

Creada en Madrid en 2011 con el fin de realizar diagnósticos genéticos desarrollando herramientas para la medicina personalizada. Cuenta con las siguientes líneas de negocio: Genotest, Farmagen y Herencia Genética Familiar.

Contiene una amplia lista de test genéticos

- **NutriTest:** Para ayudar a controlar la dieta.
- **DeporTest:** Para el rendimiento deportivo.
- **CeliaTest:** Para la predisposición a la enfermedad Celia.
- **TabacTest:** Nivel de adicción y predisposición a cáncer de pulmón.
- **AlzheimerTest:** Predisposición al padecimiento del Alzheimer.
- **TromboTest:** para la prevención de un tromboembolismo.
- **CardioTest:** Predisposición a sufrir enfermedades cardiovasculares.
- **Diabetest:** Para la predisposición a padecer diabetes mellitus tipo 2.
- **FullTest:** Estudio de 64 variaciones relacionadas con la obesidad.
- **Periodontest:** Consta de dos tipos de pruebas: **Test Microbiológico** y **Test Genético**.
- **Prueba de Parentesco (Paternidad):** Para comparar el ADN de un individuo con el del supuesto padre.
- **Prueba de Parentesco (Maternidad):** Para comparar el ADN de un individuo con el de la supuesta madre.
- **Prueba de Parentesco (Hermanidad):** Para determinar si dos individuos comparten uno o ambos progenitores.
- **Prueba de Parentesco (Abuelidad):** Para comparar el ADN de un individuo con el de la supuesto abuelo/a.
- **Prueba Genética de Sospecha:** Prueba realizada mediante la extracción del ADN de diversas fuentes y posterior comparación con el perfil de ADN de la muestra del individuo.
- **Herencia Genética Familiar:** Permite obtener información decisiva para conocer datos genéticos del individuo.



The screenshot shows the TRK GENETICS website. At the top, there is a navigation bar with the company logo and name, and links for HOME and CONTACTO. Below this is a red navigation menu with buttons for TRK GENETICS, GENOTEST, FARMAGEN, HERENCIA GENÉTICA, and LABORATORIO. The main content area features a large advertisement for TABACTEST. The ad includes a background image of a person's hands holding a cigarette, with a stethoscope visible in the background. The text in the ad describes the test as being for lung cancer predisposition and nicotine addiction, and mentions that it allows for new treatments to reduce the risk of lung cancer. At the bottom of the ad, there is a red button labeled 'COMPRAR' (BUY).

TRK GENETICS HOME | CONTACTO

TRK GENETICS GENOTEST FARMAGEN HERENCIA GENÉTICA LABORATORIO

TABACTEST. PREDISPOSICIÓN AL CÁNCER DE PULMÓN Y ADICCIÓN A LA NICOTINA.

Permite beneficiarse de nuevos tratamientos para corregir la dependencia a la nicotina, reduciendo la probabilidad de desarrollar cáncer de pulmón.

El **Servicio de Herencia Genética Familiar** consiste en la obtención de una pequeña muestra de tejido, a partir del cual el ADN se extrae y conserva en condiciones adecuadas para proceder a su análisis posterior en caso necesario como ante alguna **eventualidad clínica o legal**.

COMPRAR

National Center for Biotechnology Information NCBI

El centro nacional de información sobre biotecnología fue fundada en 1988 en Bethesda, Maryland, la cual está siendo dirigida por David Lipman y pertenece a *United States National Library of Medicine* (NLM).

Es una biblioteca de biología molecular que brinda una gran cantidad de información valiosa sobre el genoma.

Cuenta con una variedad de bases de datos que nos han sido útiles para cargar los genes y SNPs que nos interesan, en nuestro caso nos interesan las que están relacionadas con la AGA.

The screenshot shows the NCBI Gene database search results for the AR gene. At the top, there is a search bar with 'Gene' selected and 'AR' entered. Below the search bar, there are links for 'Save search', 'Limits', and 'Advanced'. The search results are displayed in a table with columns for 'Display Settings', 'Results', and 'Page'. The first result is 'AR – androgen receptor [Homo sapiens]'. A red arrow points to the 'AR' text in the first result. Below the result, there is a list of details: '1. androgen receptor', 'Official Symbol: AR', 'Other Aliases: RP11-383C12.1, AIS, DHTR, HUMARA, HYSP1, KD, NR3C4, SBMA, SMAX1, TFM', 'Other Designations: androgen nuclear receptor variant 2; dihydrotestosterone receptor; nuclear receptor', 'Location: Xq12', 'Annotation: Chromosome X, NC_000023.10 (66763874..66950461)', 'MIM: 313700', 'ID: 367', and a link 'Order cDNA clone'.

Como vimos en la imagen anterior, al buscar el gen AR en NCBI, seleccionando la base de datos **Gene**, podemos obtener algunos de los datos que necesitamos para cargar en nuestra base de datos.

Si hacemos click sobre el gen, indicado por la flecha en la imagen anterior, obtendremos más información sobre dicho gen. En este lugar podremos ver la opción “GenBank” la cual, al pincharla, nos da información sobre los nucleótidos, ARNm, los exones que codifican, etc. La base de datos GenBank contiene más de 100,000 organismos.

La base de datos de polimorfismo de nucleótido simple (**dbSNP**) complementa a **GenBank**, en esta podemos encontrar las variaciones que necesitamos. Al introducir el *rs* deseado se puede obtener el gen en el que se encuentra, la posición, la variación, etc.

NCBI Resources How To Sign in to NCBI

dbSNP SNP rs6152 Search Save search Limits Advanced Help

Display Settings: ☒ Send to: ☐

☐ rs6152 [Homo sapiens]

1. GGCAGCAGCAGCGGGAGAGCGGGGA [A/G] GCCTCGGGGGCTCCCACTTCCTCCA

MAF/MinorAlleleCount: A=0.2146/356

HGVS Names: [NC_000023.10:g.66765627G>A] [NG_009014.1:g.6754G>A] [NG_009014.2:g.6754G>A] [NM_000044.2:c.639G>A] [NP_000035.2:p.Glu213=] [NT_011669.17:g.5083615G>A] [NW_001842373.1:g.4297330G>A]

ID: 6152

MapView No VarVu PubMed GeneView SeqView No 3D OMIM

En la imagen anterior hemos buscado el rs6152 y obtuvimos información y enlaces sobre la variación. Al hacer click en el rs6152 podremos obtener más detalles del *rs* como son: la variación, el gen, el cromosoma, la posición, la citación, la notación HGVS, etc.

Si hacemos click sobre **PubMed** nos lleva a dicha base de datos.

NCBI Resources How To Sign in to NCBI

PubMed.gov PubMed rs6152 Search RSS Save search Advanced Help

US National Library of Medicine
National Institutes of Health

Display Settings: ☒ Summary, 20 per page, Sorted by Recently Added Send to: ☐

Articles that refer to rs6152

rs6152 is a A/G variation in the AR gene on human chromosome X.
[dbSNP entry](#)

Results: 8

☐ Androgen receptor genetic variants in male patients with ankylosing spondylitis in Taiwan.

1. Yu SF, Hsu YH, Cheng TT, Lai HM, Chen CJ, Kang HY.
Int J Rheum Dis. 2013 Feb;16(1):81-7. doi: 10.1111/1756-185X.12011. Epub 2012 Oct 20.
PMID: 23441776 [PubMed - in process]
[Related citations](#)

☐ Androgen receptor gene polymorphisms and risk for androgenetic alopecia: a meta-analysis.

2. Zhuo FL, Xu W, Wang L, Wu Y, Xu ZL, Zhao JY.
Clin Exp Dermatol. 2012 Mar;37(2):104-11. doi: 10.1111/j.1365-2230.2011.04186.x. Epub 2011 Oct 10. Review.
PMID: 21981665 [PubMed - indexed for MEDLINE]
[Related citations](#)

PubMed es una base de datos que brinda artículos en el área de la biomedicina. Nos muestra una serie de artículos relacionados con el gen y/o la variación que indiquemos.

Genecard

Es una base de datos genómica que contiene información de los genes y todas sus variaciones.

Está siendo desarrollada por the Crown Human Genome Center, Department of Molecular Genetics, The Weizmann Institute of Science

Ofrece enlaces a estudios publicados además de una gran cantidad de opciones de información como son: Aliases, Databases, Disorders / Diseases, Domains / Families, Drugs / Compounds, Expression, Function, Genomic Views, Intellectual Property, Orthologs, Paralogs, Pathways / Interactions, Products, Proteins, Publications, Search Box, Summaries, Transcripts y Variants.

Hemos observado que esta base de datos no está completa ya que contiene algunos genes que no tienen variaciones asociadas, y algunos de los genes que buscamos en esta base de datos no fueron encontrados.

Genecard asocia cada una de sus variaciones con *Ensembl* y con *NCBI (dbSNP)* para mostrar más detalles sobre sus efectos fenotípicos, cromosoma y posición en el que se encuentra, etc.

10/1928 NCBI SNPs in [AR](#) are shown (see all 1928) [improved](#) [About this table](#)

Genomic Data					Transcription Related Data				Allele Frequencies				
SNP ID	Valid	Clinical significance	Chr X pos	Sequence	#	AA Chg	Type	More	#	Allele freq	Pop	Total sample	More
<div>-</div>	<div>1st</div>	<div>2nd</div>	<div>3rd</div>	--	--		<div>-</div>	--	--	<div>-</div>	<div>-</div>	<div>-</div>	<div>Sort</div>
rs104894742 ^{1.2}	C	pathogenic	66764992(+)	GGATGA/GAAGTG	2	K E	mis ¹		0	--	--	--	--
rs1800053 ^{1.2}	C,F	pathogenic	66931295(+)	AGAGGC/ATTCCA	4	/D /A	mis ¹		1		NA	4320	
rs137852593 ^{1.2}	C	pathogenic	66937326(+)	CTTCCG/TCAACT	4	R L	mis ¹		0	--	--	--	--
rs137852562 ^{1.2}	C	pathogenic	66941679(+)	GGTACC/TGCATG	4	R C	mis ¹		0	--	--	--	--
rs137852591 ^{1.2}	C,F	pathogenic	66941751(+)	GGCTCC/GAAATC	4	Q E	mis ¹		1		NA	4552	
rs137852577 ^{1.2}	C	pathogenic	66942740(+)	TCGATC/TGTATC	4	R C	mis ¹		0	--	--	--	--
rs9332969 ^{1.2}	C,F,H	pathogenic	66942741(+)	CGATCG/ATATCA	4	/H /R	mis ¹ ese ³		4		EA NS NA	264	
rs9332971 ^{1.2}	C,F,H	pathogenic	66942786(+)	AAGACG/ACTTCT	4	/H /R	mis ¹ ese ³		6		EA NS NA	506	
rs137852600 ^{1.2}	C	probable-pathogenic	66937377(+)	GATGGA/G/TGCTCA	6	E G V	mis ¹		0	--	--	--	--
rs137852564 ^{1.2}	C	probable-pathogenic	66942818(+)	ACTCCA/G/TTGCAG	6	M V L	mis ¹		0	--	--	--	--

Ensembl

En este proyecto trabajan en conjunto *The European Bioinformatics Institute* y el *Wellcome Trust Sanger Institute* en el cual desarrollan un software que produce anotaciones automáticas en los genomas eucariotas. Este proyecto se inicio en 1999.

6,765,127-66,766,127 Variation: rs6152

rs6152 SNP

Original source

Alleles

Location

Co-located

Evidence status

Synonyms

HGVS names

Genotyping chips

Explore this variation

Variants (including SNPs and indels) imported from dbSNP (re
Reference/Alternative: **G/A** | Ancestral: G | Ambiguity code: R
Chromosome X:66765627 (forward strand) | [View in location ta](#)
with HGMD-PUBLIC [CM051843](#)

Archive dbSNP [rs1800052](#)

This variation has 6 HGVS names - click the plus to show

This variation has assays on: Illumina_HumanOmni2.5

Genomic context

Genes and regulation

Population genetics

Individual genotypes

Linkage disequilibrium

Phenotype data

Citations

Phylogenetic context

Flanking sequence

ATTGATT
CGGSGTG
TCATGCT

En la imagen anterior se puede observar los detalles de cada variación, su efecto fenotípico, citas, etc....

En **Ensembl** no todas las variaciones encontradas están relacionadas a un fenotipo. Apoyan sus variaciones en papers para poder dar una mayor fiabilidad con respecto a los rasgos fenotípicos ocasionados por dichas variaciones.

En la siguiente tabla se muestran las variaciones que Ensembl utiliza con respecto a la Alopecia Androgenética.

Nombre	Gen	RS
Intergenic	Intergenic	rs2180439
Intergenic	Intergenic	rs201571
Paired Box Gene 1, Paired Box Protein Pax -1	PAX1	rs1160312
Androgen Receptor	AR	rs6152

Variaciones de Ensembl [3] [10] [11]

SNPedia

Es una wiki en la cual podemos encontrar información relevante acerca de los SNPs (Single Nucleotide Polymorphism).

Fue fundada por Greg Lennon y Mike Carias, genetista y programador respectivamente.

En esta wiki podemos encontrar el o los genes y cromosoma relacionados con un snp en concreto; enlaces a papers publicados que estudian el snp en cuestión; una descripción y enlaces a distintas bases de datos genómicas que utilizan y hablan de dicho snp como son: openSNP, Ensembl, dbSNP entre otras.

SNPedia Muchas veces asocia los SNPs a alguna enfermedad o efecto fenotípico.

Hay que destacar que esta wiki, a la fecha de hoy, no tiene todos los SNPs del genoma registrado aunque al paso del tiempo ha ido incrementando su información gradualmente.



Metodología SILE

Es una serie de pasos que nos permite obtener toda la información necesaria y relevante sobre una enfermedad genética, ya sea una mutación o una variación. Seguir estos pasos nos permite tener un trabajo correcto.

Esta consiste en cuatro esenciales:

✓ Search

En este paso buscamos toda la información que de alguna manera tiene que ver con la enfermedad genética en cuestión.

La información que en este paso obtenemos consiste en: genes, variaciones, cromosoma, bases de datos genómicas, empresas de test genéticos, enfermedades genéticas, papers, entre otras.

El volumen de información aquí obtenida es considerable.

✓ Identification

Se filtra toda la información obtenida en el paso *Search* para quedarnos solo con aquellas variaciones o mutaciones que están profundamente relacionadas con una enfermedad, esto se hace mediante artículos publicados en los cuales se realizaron estudios en diferentes poblaciones y con la ayuda de biólogos y/o genetistas.

También se identifican las bases de datos genómicas más importantes.

✓ Load

Se realiza la arquitectura de la base de datos, se carga toda la información que se requiere en la BD. La información que se carga en este paso son las variaciones, genes y demás que hemos obtenido en el paso de *Identificación*.

✓ Exploitation

Se desarrollan todos los contenidos que se cargarán en la web, estos contenidos consistirán en texto, imágenes, entre otros. Este proceso se va desarrollando desde el principio.

También se especifica la empresa en la que se explotaran los contenidos

Esta metodología es la que estamos utilizando en el desarrollo de este trabajo, ya que nos brinda una mayor facilidad para obtener información correcta y fiable para poder cumplir con las necesidades.

4. Solución del problema

4.1 Especificación de genes y variaciones relacionadas con la AGA.

Hemos recopilado información de distintas bases de datos genómicas como son: ncbi, genecard, ensembl y snpedia. Obtuvimos todos los genes implicados en el padecimiento de la Alopecia Androgenética y las variaciones encontradas en cada gen.

En la búsqueda que realizamos encontramos los genes: AR, EDA2R, HDAC9, HDAC4, TARDBP, C1orf127-TARDBP, PAX1, FOXA2, MAPT-AS1, SPPL2C, SETBP1, SLC14A2 cerca del gen SETBP1 y AUTS2; y todas las variaciones asociadas a estos genes. También encontramos variaciones que no están asociadas a ningún gen y que se han denominado como INTERGENIC.

En la siguiente tabla presentamos la lista de genes y variaciones que hemos encontrado relacionadas con la AGA.

Nombre	Gen	RS
Androgen Receptor	AR	rs6152
		rs2497938
		rs6625163
Ectodysplasin A2 Receptor	EDA2R	rs1385699
Histone Deacetylase 9	HDAC9	rs756853
		rs2073963
		rs2249817
Intergénica	Intergenic	rs2223841
		rs201571
		rs2180439
		rs6137444
		rs1998076
		rs6113491
		rs10521339
		rs938059
	EDA2R/AR Intergenic Region	rs12558842
Histone Deacetylase 4	HDAC4	rs9287638
TAR DNA-Binding Protein	TARDBP	rs12565727
Chromosome 1 Open Reading Frame 127	C1orf127 –TARDBP	rs2003046
Paired Box Gene 1, Paired Box Protein Pax -1	PAX1	rs1160312

FOXA2: FORKHEAD BOX A2	PAX1, FOXA2	rs6047844
AUTISM SUSCEPTIBILITY CANDIDATE 2	AUTS2 También conocido como: FBRSL2	rs6945541
SPPL2C: Signal Peptide Peptidase Like 2C MAPT: Antisense RNA 1	17q21.31 (MAPT-AS1, SPPL2C)	rs12373124
SETBP1: Set Binding Protein 1	SETBP1	rs10502861
Solute Carrier Family 14 (Urea Transporter), Member 2	SLC14A2 cerca del gen SETBP1	rs8085664

No todos los genes tienen la misma importancia y no todos se han estudiados a fondo; esto nos obliga filtrar la información encontrada para utilizar solo los genes y variaciones que están más estudiados y validados por biólogos y genetistas.

Para poder filtrar todos los genes y variaciones obtenidas, hicimos uso de artículos en los cuales se hicieron estudios, sobre la AGA, en determinadas poblaciones. En cada estudio se tomó una cantidad de personas considerable como *control* (*sanos*) y otra cantidad como *case* (*enfermos*).

En estos estudios se pudo observar que algunas variaciones incrementan considerablemente la predisposición a padecer la AGA, mientras que otras aumentan la posibilidad de padecerla a una edad temprana; además se determinaron genes que son más comunes en determinadas poblaciones ya que, genéticamente, una población puede ser diferente de otra.

Los SNPs que hemos identificado por ser los que presentan un mayor nivel de riesgo son los siguientes [9].

SNP	Reference / Alternative	Odds Ratio	P.Value	Bloque
rs12565727	A/G	1.33 (1.22–1.45)	9.07×10^{-11}	Otros
rs9287638	C/A	1.31 (1.21–1.41)	1.01×10^{-12}	Otros
rs2073963	T/G	1.29 (1.20–1.38)	1.08×10^{-12}	HDAC9
rs6945541	C/T	1.27 (1.18–1.38)	1.71×10^{-9}	Otros
rs12373124	T/G/A/C	1.33 (1.21–1.45)	5.07×10^{-10}	Otros
rs10502861	C/T	1.28 (1.18–1.39)	2.62×10^{-9}	Otros
rs6047844	T/C	1.60 (1.49–1.72)	1.71×10^{-39}	PAX1/FOXA2
rs2497938	T/C	2.20 (2.04–2.37)	2.40×10^{-91}	AR/EDA2R

rs12565727

Está situado en el cromosoma 1, en la posición 11033082. Esta posición se corresponde con uno de los intrones del gen C1orf127, codificante y validado por NCBI pero de función aún desconocida.

Además, se sitúa a 39 kb en 5' del gen TARDBP, que codifica para la proteína de unión a TAR (esta proteína tiene la función de unirse a las posibles secuencias TAR insertadas por el virus del VIH en el genoma, e impedir el reconocimiento de dichas secuencias por el factor de transcripción vírico Tat, evitando así la replicación vírica). Otra función que se le atribuye es regular el splicing del mRNA del receptor transmembrana CFTR (involucrado en fibrosis quística), y estabilizar la forma corta del mRNA codificante para proteínas de neurofilamentos (GeneCards).

rs9287638

Se sitúa en la posición 239694631 del cromosoma 2. Es intergénico.

Su entorno genético es el siguiente:

- Está a unas 276 kb en 3' respecto del gen HDAC4
- Está a unas 61 kb en 3' respecto del gen codificante de proteínas LOC100287387 (aún sin función conocida).
- Está a unas 60 kb en 5' respecto del gen TWIST2 (factor de transcripción, se cree que tiene importancia en mantener inhibida la diferenciación de osteoblastos durante el desarrollo).

Hasta el momento, la explicación biológica más aceptada es similar a la que se aplica a HDAC9: se trata de histona desacetilasas que modifican la expresión de otros genes. Sin embargo, recordemos que el SNP propiamente dicho está algo lejano del gen HDAC4.

En población europea los homocigotos para el alelo con A de este SNP tienen 1.31 veces más posibilidades de sufrir alopecia androgénica.

rs2073963

Situado en la posición 18877874 del cromosoma 7, se trata de un cambio de G por T en un intrón de HDAC9. Si bien este SNP muestra asociación con el rasgo, las investigaciones parecen indicar que se trata de un efecto debido al desequilibrio de ligamiento con rs75683.

rs6945541

Es un SNP intergénico, que se ubica en el cromosoma 7, al igual que el bloque HDAC9. Sin embargo, está a varias decenas de megabases de éste (descartado por tanto que forme haplotipos habitualmente con los SNPs de ese bloque).

El gen más cercano es AUTS2 (autism susceptibility candidate 2), codificante para proteína. El SNP se encuentra a 460 kb en 5' respecto al gen.

La relación funcional entre el rasgo y el gen AUTS2 está por descubrir, pero se sabe que este gen se expresa abundantemente en el tejido del cabello, en cuero cabelludo, y en la piel, mientras que en otros tejidos como la sangre está ausente.

El estudio (Li et al. 2012) lo denota como el SNP más relevante de este locus (P-value del orden de 10^{-9}).

rs12373124

Cambio de T por A situado en la posición 43924219 del cromosoma 17, correspondiente a dos genes.

- Corresponde a un exón del gen SPPL2C (*Signal peptide peptidase like 2C*), también llamado IMP5 (*intramembrane peptidase 5*), codificante para proteína. El artículo (Li et al, 2012), lo presentan como un SNP sinónimo que da lugar al cambio His649His.
- También corresponde a un intrón del gen MAPT-AS1 (*MAPT antisense RNA 1*), no codificante para proteína, donde se sitúa lejos de los puntos de *splicing*.

En el artículo (Li et al. 2012), lo presentan como el SNP más relevante de varios encontrados en esta región del genoma, con un P-value del orden de 10^{-10} . Los autores advierten que se han encontrado 118 SNPs (significativos a nivel del genoma completo) en esta región, y que probablemente no sea ésta la variación que cause realmente el fenotipo, sino una en elevado desequilibrio de ligamiento con ésta.

rs10502861

Situado en el cromosoma 18, es un cambio de T por C en la posición 42800148, correspondiente a un intrón del gen SLC14A2 (*solute carrier family 14 (urea transporter), member 2*), en una región alejada de puntos de *splicing*.

En el estudio (Li et al. 2012), se presenta el SNP como el principal de un bloque de 16, con un P-value de 10^{-9} . Según ellos, la importancia podría ser más por situarse en 3' del gen SETBP1 (SET binding protein 1), que por el gen en el que se encuentra, ya que éste segundo gen sí se expresa en los tejidos de nuestro interés (pelo, piel, cuero cabelludo), pero no en células sanguíneas. Según NCBI, la proteína de este gen se transloca al núcleo y actúa uniéndose al oncogén SET, involucrado en la replicación del DNA.

rs6047844

Está situado en la posición 22037575 del **cromosoma 20**, entre PAX1 y FOXA2. El estudio (Li et al. 2012) lo presenta como el SNP más importante del bloque, con un P-value conjunto (para todo el experimento) en ese SNP del orden de 10^{-39} . En el estudio se indica que hay otros 277 SNPs con asociación significativa con el trait, confirmando la relevancia de esta región en la alopecia androgénica.

rs2497938

Cambio de C por T en la posición 66563018. Es un SNP situado en el cromosoma X, en la región entre los genes AR y EDA2R. En esta región hay pocos genes.

El estudio (F F Brockschmidt et al. 2011) lo presenta como el SNP más relevante de este locus (P-value = $3,11E(-22)$, odds ratio (OR) 6,50), en una muestra de población alemana validada más tarde con resultados de población australiana caucásica.

4.2 Desarrollo del proceso de probabilidades de riesgo.

La probabilidad que tiene un SNP para manifestar el efecto fenotípico de una enfermedad en concreto, no está directamente disponible en ninguno de los papers estudiados.

Aunque estos papers no nos dan una probabilidad para cada SNP, podemos encontrar en ellos unos datos que pueden ayudarnos a determinar estas probabilidades. Estos datos son: el **odds ratio**, el **p.value** y el **etiologic fraction**.

Entre ellos, el que más nos interesa es el Odds Ratio (OR).

Odds Ratio (OR)

Es una medida estadística que nos brinda la posibilidad de que una enfermedad se manifieste en las personas en presencia de un riesgo frente a las personas que no están en presencia de ese riesgo.

Un OR de 2.34 de un alelo indica que las personas con ese alelo tienen 2.34 veces más riesgo de presentar la enfermedad que aquellos que no tienen el alelo.

Para hacer el calcula del OR utilizamos una tabla 2x2 en la cual identificamos los casos (enfermos) y los controles (sanos) frente al alelo de riesgo y de referencia.

El siguiente caso es hipotético en el cual se estudian las personas con AGA.

	Case	Control
Alelo de riesgo	645	195
Alelo de referencia	355	805

Se obtiene el cociente del total de personas que están enfermas y presentan el alelo de riesgo, y el total de personas que están enfermas y no presentan el alelo de riesgo; También obtenemos el cociente del total de personas sanas con el alelo de riesgo, y el total de personas sanas sin el alelo de riesgo.

Una vez obtenido estos dos cocientes, se divide el primero entre el segundo y el resultado sería el **Odds Ratio**. A continuación mostramos la fórmula para el cálculo del OR

$$\text{Odds Ratio (OR)} = \frac{\text{Case (alelo de riesgo/alelo de referencia)}}{\text{Control (alelo de riesgo/alelo de referencia)}}$$

El ejemplo anterior se resolvería de la siguiente manera:

$$\frac{645/355}{195/805} = \frac{645 * 805}{195 * 355} = \frac{519225}{69225} = 7.50$$

Vemos que el alelo en cuestión muestra una relación con la enfermedad. El resultado final dice que el alelo de riesgo esta 7.5 veces más en los enfermos que en los sanos.

Esto expresa la proporción de veces que la Alopecia Androgenética aparece en presencia del alelo de riesgo frente al alelo de referencia.

Probabilidad

Transformaremos el OR en probabilidades, para ello utilizamos la siguiente formula.

$$P = \frac{OR}{OR+1}$$

$$P = \frac{7.5}{7.5+1} = \frac{7.5}{8.5} = 0.882 \rightarrow 88.2\%$$

La probabilidad de que la Alopecia Androgenética ocurra en presencia del alelo es del 88.2%.

A continuación mostramos los porcentajes de probabilidad de cada una de las variaciones que identificamos en la manifestación de la Alopecia Androgenética. Estas probabilidades la obtuvimos al aplicar la formula anteriormente explicada.

SNP	OR	CI	Probabilidad
rs12565727	1.33	95% (1.22 – 1.45)	57.08%
rs9287638	1.31	95% (1.21 – 1.41)	56.71%
rs2073963	1.29	95% (1.20 – 1.38)	56.33%
rs6945541	1.27	95% (1.18 – 1.38)	55.95%
rs12373124	1.33	95% (1.21 – 1.45)	57.08%
rs10502861	1.28	95% (1.18 – 1.39)	56.14%
rs6047844	1.60	95% (1.49 – 1.72)	61.54%
rs2497938	2.20	95% (2.04 – 2.37)	68.75%

Cuando en el código genético de una persona se detecta una de estas variaciones, el nivel de probabilidad de riesgo será el que corresponde con dicha variación.

Intervalo de Confianza (IC)

Es también llamado CI por sus siglas en ingles. Es el rango en que se encuentra el valor verdadero del **Odds Ratio**, con una probabilidad de acierto.

Para poder determinar el intervalo de confianza tenemos que utilizar la siguiente formula.

$$IC = OR^{(1 \pm z / x)}$$

Z: identifica una constante que depende del porcentaje. La constante del 95% es 1.96.

X: Se obtiene con el Chi cuadrado.

Para calcular el Chi cuadrado se utiliza la siguiente formula.

$$X^2 = \frac{(t - 1)(a * d - b * c)^2}{n1 * n0 * m1 * m0}$$

Para poder obtener el intervalo de confianza lo primero que se hace es calcular el OR. Los datos para hacer estos cálculos están definidos en la siguiente tabla.

	Case	Control	Totales
Alelo de riesgo	a 575	b 162	m1 737
Alelo de referencia	c 425	d 838	m0 1263
Totales	n1 1000	n0 1000	T 2000

$$OR = \frac{575/425}{162/838} = \frac{575 * 838}{162 * 425} = \frac{481850}{68850} = 7$$

Utilizando la formula anteriormente mostrada obtenemos el valor de X

$$X^2 = \frac{(2000 - 1)(575 * 838 - 162 * 425)^2}{1000 * 1000 * 737 * 1263} = \frac{340967431000000}{930831000000}$$

$$X = \sqrt{366.304335588} = 19.1391$$

Ahora podemos calcular el intervalo de confianza, el cual tiene un límite superior y un límite inferior.

Límite superior

$$IC = 7^{(1+1.96/19.1391)} = 7^{(1.10240815921)}$$

$$IC = 8.5436$$

Límite inferior

$$IC = 7^{(1-1.96/19.1391)} = 7^{(0.89759184078)}$$

$$IC = 5.7352$$

Llegado a este punto hemos visto que un OR de 7 se encuentra en un intervalo de confianza de 8.5436 y 5.7252.

Los límites del intervalo de confianza deben estar por encima de la unidad para que el OR se pueda considerar como significativo.

4.3 Ampliación del modelo conceptual

4.3.1 Modelado de las probabilidades de riesgo

Debido a la importancia que tiene el hecho de identificar las probabilidades de riesgo de los SNPs para que una persona padezca la Alopecia Androgénica u otra enfermedad genética, hemos tenido que modelar todo el proceso relacionado con las probabilidades de riesgo. Este modelado lo haremos sobre el modelo conceptual desarrollado en el centro de investigación en métodos de producción de software (PROS) de la UPV.

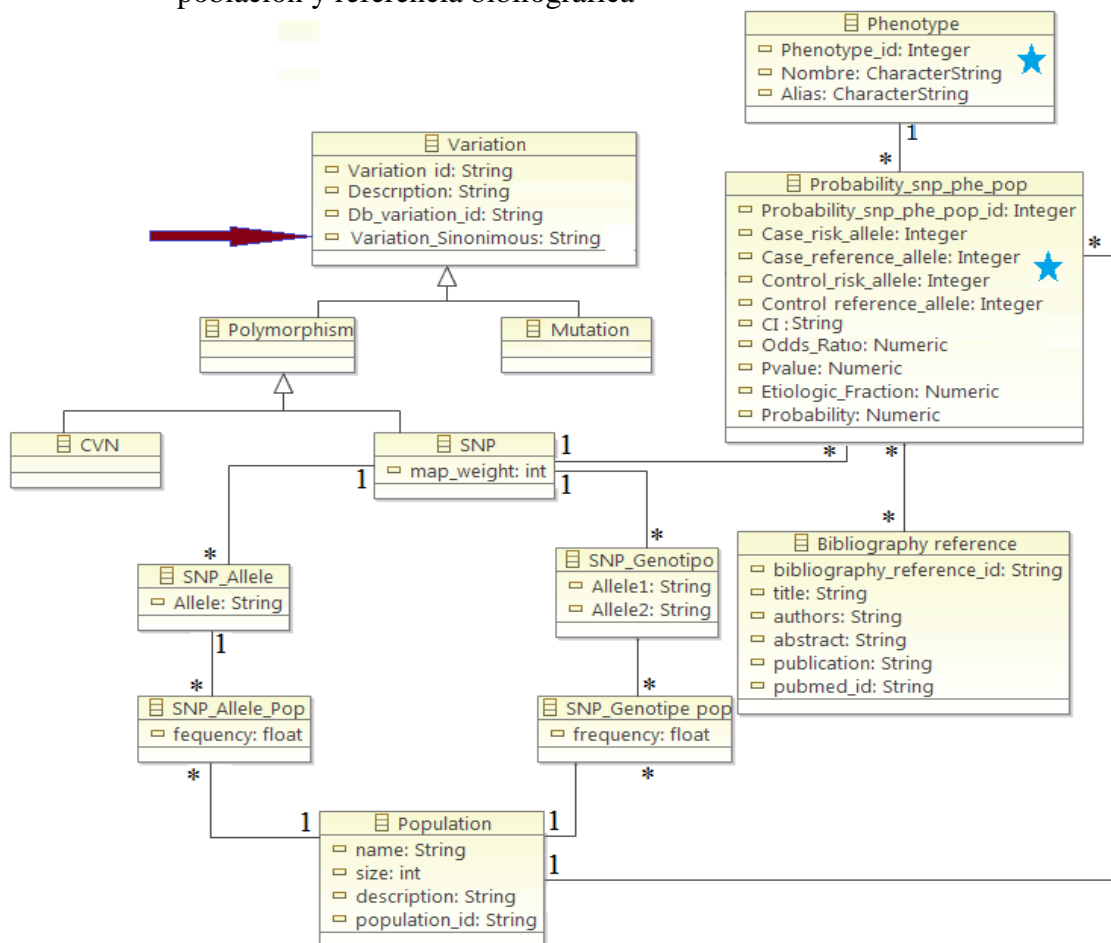
Se necesita que el modelo conceptual pueda manipular las probabilidades permitiendo identificar el odds ratio, el p.value y el etiologic fraction.

El modelo conceptual debe cumplir con lo siguiente:

1. Un SNP puede afectar a más de un fenotipo y tener probabilidades iguales o distintas para cada uno.
2. Cada probabilidad está asociada directamente a una población.
3. Una probabilidad puede estar dada por muchas referencias bibliográficas y viceversa.

A modo de restricción

1. Un SNP no puede tener más de una probabilidad con el mismo fenotipo, población y referencia bibliográfica



Hemos añadido al modelo conceptual algunas clases y sus relaciones.

Clases añadidas

1. Probability_snp_phe_pop
2. Phenotype

Relaciones

1. Un SNP puede tener muchas probabilidades.
2. Una probabilidad pertenece a un único SNP.
3. Una probabilidad se relaciona con un único fenotipo.
4. Un fenotipo puede tener muchas probabilidades.
5. Una probabilidad está asociada a una única población
6. Una población puede tener muchas probabilidades.
7. Una probabilidad puede estar relacionada con muchas referencias bibliográficas
8. Una referencia bibliográfica puede contener muchas probabilidades.

La clase **Probability_snp_phe_pop** indica la probabilidad que tiene un SNP para manifestar un fenotipo respecto a una población, el cual está indicado en un estudio determinado que lo identificamos en la clase de referencias bibliográficas.

Esta clase contiene una serie de atributos que deben considerarse a la hora de determinar la probabilidad.

Probability_snp_phe_pop	
Probability_snp_phe_pop_id:	Integer
Case_risk_allele:	Integer
Case_reference_allele:	Integer
Control_risk_allele:	Integer
Control_reference_allele:	Integer
CI :	String
Odds_Ratio:	Numeric
Pvalue:	Numeric
Etiologic_Fraction:	Numeric
Probability:	Numeric

1. **Probability_snp_phe_pop_id:** es un identificador auto-incrementable
2. **Case_risk_allele:** Identifica la cantidad de personas enfermas con el alelo de riesgo utilizado en un estudio determinado
3. **Case_reference_allele:** La cantidad de personas enfermas con el alelo de referencia utilizado en el estudio
4. **Control_risk_allele:** Identifica la cantidad de personas sanas con el alelo de riesgo utilizado en el estudio.
5. **Control_reference_allele:** Las personas sanas con el alelo de referencia utilizado en el estudio.
6. **CI:** Es el rango en que se encuentra el verdadero valor del Odds Ratio con una probabilidad de acierto.
7. **Odds_Ratio, Pvalue y Etiologic_Fraction** son cálculos estadísticos que se utilizan en los estudios para determinar distintos tipos de probabilidad.
8. **Probability:** para determinar la probabilidad de que se manifieste el fenotipo, esta es obtenida a través del odds ratio.

La clase **Phenotype** identifica las enfermedades genéticas.

1. **Phenotype_id:** es un identificador que se incrementa automáticamente.
2. **Nombre:** Es el nombre de la enfermedad.
3. **Alias:** Para todos los nombres alternativos de una enfermedad. Si un fenotipo tiene más de un nombre alternativo, se pone separado por una coma.

Phenotype	
Phenotype_id: Integer	★
Nombre: CharacterString	
Alias: CharacterString	

4.3.2 Soporte de sinónimos de RS en el modelo conceptual

Existen SNPs que se identificaron con distintos nombres como el **rs201571** que también puede encontrarse con el nombre rs582426 y rs57456154. Por esta razón hemos añadido el atributo Variation_Sinonimous de tipo String que nos permitirá introducir todos los alias de una variación separados por coma.

Variation	
Variation id: String	★
Description: String	
Db_variation_id: String	
Variation_Sinonimous: String	

4.4 Carga de variaciones

Para poder cargar las variaciones hemos tenido que ampliar la base de datos, desarrollada por el centro de investigación en métodos de producción de software Pros a partir del modelo conceptual, para que se adecue a los cambios que hemos hecho, de manera que podamos introducir y administrar todos los datos relacionados con las probabilidades.

A continuación veremos la estructura de cada una de las tablas que hemos ampliado o agregado en el esquema de bases de datos. Todas las tablas, el esquema de base de datos y las restricciones la hemos creado con Microsoft SQL SERVER 2008¹¹.

Probability_snp_phe_pop

	Column Name	Data Type	Allow Nulls
?	Probability_snp_phe_pop...	int	<input type="checkbox"/>
	Case_risk_allele	int	<input checked="" type="checkbox"/>
	Case_reference_allele	int	<input checked="" type="checkbox"/>
	Control_risk_allele	int	<input checked="" type="checkbox"/>
	Control_reference_allele	int	<input checked="" type="checkbox"/>
	odds_Ratio	numeric(18, 2)	<input type="checkbox"/>
	CI	nvarchar(50)	<input checked="" type="checkbox"/>
	PValue	numeric(18, 2)	<input checked="" type="checkbox"/>
	Etiologic_fraction	numeric(18, 2)	<input checked="" type="checkbox"/>
▶	Probability	numeric(18, 2)	<input checked="" type="checkbox"/>
	Phenotype_id	int	<input checked="" type="checkbox"/>
	SNP_id	int	<input type="checkbox"/>
	Population_id	nvarchar(20)	<input checked="" type="checkbox"/>

Esta es la tabla más importante para el desarrollo de las probabilidades, ya que en ella guardamos todos los datos y valores que se utilizan para el cálculo de la probabilidad que un SNP tiene para la manifestación de un efecto fenotípico.

Los campos Phenotype_id, SNP_id y Population_id son claves ajenas de las tablas Phenotype, SNP y Population respectivamente.

A continuación creamos un trigger¹² para calcular la probabilidad automáticamente al momento de hacer un insert o un update en esta tabla.

Para esto utilizamos la formula:
$$P = \frac{OR}{OR+1}$$

¹¹ Es un sistema gestor de bases de datos desarrollado por Microsoft



¹² También conocido como disparador o desencadenador, es un procedimiento almacenado que se ejecuta al cumplirse una condición especificada. Esta condición puede ser un insert (inserción), un update (actualización) y/o un delete (borrado).


```

CREATE TRIGGER [dbo].[tr_Probability_Calculation]
ON [dbo].[Probability_snp_phe_pop]
AFTER INSERT, UPDATE
AS
begin
--Se calcula y se actualiza la probabilidad a partir del odds ratio de la variacion.
update Probability_snp_phe_pop set Probability = (Select(Odds_Ratio
/(odds_Ratio + 1) * 100) from inserted)
where Probability_snp_phe_pop_id = (select
Probability_snp_phe_pop_id from inserted)
end

```

Variation

	Column Name	Data Type	Allow Nulls
	Variation_id	int	<input type="checkbox"/>
	chr_gene_id	int	<input checked="" type="checkbox"/>
	Db_version_id	int	<input checked="" type="checkbox"/>
	Description	nvarchar(50)	<input checked="" type="checkbox"/>
	Db_variation_id	nvarchar(50)	<input checked="" type="checkbox"/>
	Variation_Sinonimous	nvarchar(MAX)	<input checked="" type="checkbox"/>
	Clinically_important	nvarchar(50)	<input checked="" type="checkbox"/>
	Private	bit	<input checked="" type="checkbox"/>
	NC_indentifier	nvarchar(50)	<input checked="" type="checkbox"/>
	Frequency_variation_type	bit	<input type="checkbox"/>

Esta tabla se utiliza para introducir todas las variaciones. La hemos ampliado agregando los campos Variation_Sinonimous y Frequency_variation_type.

Variation_Sinonimous: Lista de sinónimos, separados por comas, que puede tener una variación.

Frequency_variation_type: En el modelo conceptual las variaciones son especializadas en Polymorphism y Mutation; debido a que estas especializaciones no tienen atributo, hemos definido el campo Frequency_variation_type de tipo “**bit**”¹³ en la base de datos para identificar cuando una variación es un polimorfismo o una mutación; es un campo booleano que recibe como valor true o false, cuando sea un polimorfismo habrá un true y cuando sea una mutación habrá un false. Este campo es obligatorio.

¹³ Este tipo de datos actúa como booleano

Population

	Column Name	Data Type	Allow Nulls
🔑	Population_id	nvarchar(20)	<input type="checkbox"/>
	Name	nvarchar(30)	<input checked="" type="checkbox"/>
	size	int	<input checked="" type="checkbox"/>
	Description	nvarchar(MAX)	<input checked="" type="checkbox"/>

Esta tabla se utiliza para poner todas las poblaciones en las que se realizan estudios genéticos sobre la relación genotipo-fenotipo.

Sus campos:

Population_id: es el identificador único

Name: nombre de la población

Size: representa el tamaño total de la población

Description: Una breve descripción sobre la población

El modelo conceptual utiliza las poblaciones pero aun no estaba especificado en el esquema de base de datos. Para trabajar con las probabilidades es necesario especificar las poblaciones, ya que las probabilidades dependen directamente de estas.

Reference_Variation

	Column Name	Data Type	Allow Nulls
🔑	Probability_snp_phe_pop...	int	<input type="checkbox"/>
🔑	Bib_ref_id	int	<input type="checkbox"/>

Esta nos permite hacer una relación entre la probabilidad que tiene cada SNP y su referencia bibliográfica.


Una misma probabilidad no puede tener la referencia bibliográfica repetida, pero si puede tener más de una referencia bibliográfica; por este motivo la clave primaria de esta tabla es la combinación de ambos atributos impidiendo que se repitan las distintas combinaciones que pueden resultar de estos campos.

Sus campos:

Probability_snp_phe_pop_id: Es una clave ajena de la tabla Probability_snp_phe_pop

Bib_ref_id: Es una clave ajena de la tabla Bibliography_reference

SNP

	Column Name	Data Type	Allow Nulls
	Variation_id	int	<input type="checkbox"/>
	Map_weight	int	<input type="checkbox"/>


Aquí se identifican todas las variaciones que son de tipo SNP.

Sus campos

Variation_id: Es la clave primaria de esta tabla y también una clave ajena de la tabla Variation.

Map_weight: Identifica las veces que se ha mapeado un SNP en el genoma de un individuo. Este campo es obligatorio.

CVN


	Column Name	Data Type	Allow Nulls
	Variation_id	int	<input type="checkbox"/>
	Repetition	int	<input type="checkbox"/>

Aquí se identifican todas las variaciones que son de tipo CVN

Variation_id: Es la clave primaria de esta tabla y también una clave ajena de la tabla Variation.

Repetition: Cantidad de veces que se repite o se borra una región de la secuencia de ADN. Este campo es obligatorio.

Phenotype

	Column Name	Data Type	Allow Nulls
	Phenotype_id	int	<input type="checkbox"/>
	Name	nvarchar(50)	<input type="checkbox"/>
	Alias	nvarchar(80)	<input checked="" type="checkbox"/>

Para las enfermedades genéticas que tienen relación con ciertos SNPs.

Sus campos:

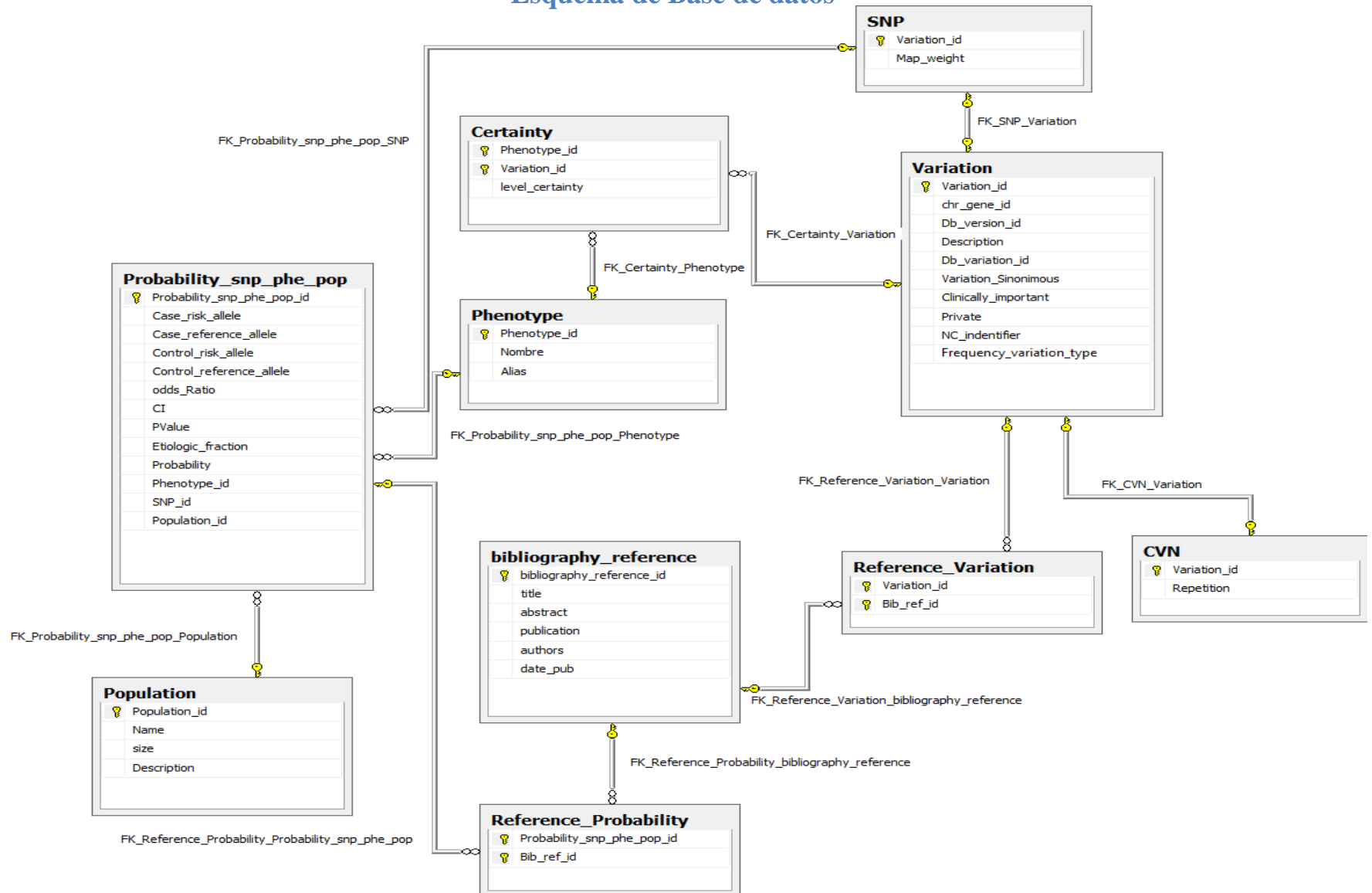
Phenotype_id: Es el identificador único.

Name: Representa el nombre de la enfermedad. Este campo es obligatorio.

Alias: Otros nombre por los que también se conoce la enfermedad.

Las clases SNP y CVN están especificadas en el modelo conceptual pero aun no estaban en el esquema de base de datos, de manera que lo hemos añadido para que el esquema de base de datos esté mas completo.

Esquema de Base de datos



El esquema de base de datos anterior fue desarrollado en Microsoft SQL Server 2008, en dicho esquema se puede observar que una variación puede ser de tipo polimorfismo o mutación indicada en el campo Frequency_variation_type.¹⁴ Los polimorfismos pueden ser del tipo SNP o del tipo CVN.

La tabla Probability_snp_phe_pop se relaciona con las tablas SNP, Phenotype y Population a través de los campos SNP_id, Phenotype_id y Population_id los cuales actúan como claves ajenas. También existe una relación de mucho a mucho con la tabla Bibliography_reference; para lograr esta relación, tuvimos que crear una tabla intermedia llamada Reference_Probability la cual tiene como clave primaria la combinación de dos campos que también actúan como claves ajenas de las tablas Probability_snp_phe_pop y la tabla Bibliography_reference¹⁵.

Este esquema también cumple con lo siguiente:

- Un SNP puede afectar a más de un fenotipo y tener probabilidades iguales o distintas para cada uno.
- Cada probabilidad está asociada directamente a una población.
- La probabilidad de un SNP puede estar dada por muchas referencias bibliográficas
- Una referencia bibliográfica puede proveer muchas probabilidades

El esquema anterior aun está incompleto, ya que falta especificarles las restricciones correspondientes para que cumpla con lo siguiente:

1. Un SNP no puede tener dos probabilidades o más con las mismas características.
 - Esto impide que un SNP tenga el mismo fenotipo, población y referencia bibliográfica varias veces, ya que ningún paper estudia más varias veces un mismo SNP respecto a una población y un fenotipo.
2. Las variaciones de tipo SNP y las variaciones de tipo CVN deben estar definida como polimorfismo.
 - En el modelo conceptual las variaciones especializadas como polimorfismo se especializan a su vez como SNP o CVN.
3. Las variaciones no pueden ser SNP y CVN al mismo tiempo.
4. Los valores de los casos y los controles en una probabilidad deben ser mayor a cero.

Además también crearemos algunos procedimientos almacenados¹⁶ que permitirán a la base de datos calcular el OR y el IC en el momento necesario.

¹⁴ Este campo es de tipo Boolean, True indica que es un polimorfismo y False indica que es una mutación.

¹⁵ Los papers que hemos utilizado para obtener y estudiar los SNPs.

¹⁶ Es un conjunto de instrucciones almacenadas y ejecutadas por el motor de base de datos.

Restricciones para la base de datos

- Un SNP no puede tener dos probabilidades o más con las mismas características.

La siguiente restricción es un trigger que se especifica en la tabla **Reference_Probability**, y se ejecutará al momento de hacer un insert o un update.

```
CREATE TRIGGER [dbo].[tr_Probability_Restriction] (Nombre del trigger)
ON [dbo].[Reference_Probability] (Tabla en la que se ejecuta)
AFTER INSERT, UPDATE (Evento por el cual se ejecuta el trigger)
AS
begin (Inicio de las instrucciones)
    --Variables.
    declare @phe int;
    declare @s int;
    declare @pop int;
    declare @bib int;
    declare @cont int;
    --Se llenan las variables con los datos deseados a partir de la insercion o la
    --actualizacion que se haya hecho.
    set @bib = (select Bib_ref_id from inserted17)

    Select @phe = Phenotype_id, @s = SNP_id, @pop = Population_ID from
        Probability_snp_phe_pop
    where Probability_snp_phe_pop.Probability_snp_phe_pop_id = (Select
        Probability_snp_phe_pop_id from inserted)
    --Se cuentan los registros de probabilidad que tienen las mismas características que
    --que el que se esta utilizando.
    set @cont = (select count(*) from Probability_snp_phe_pop join
        Reference_Probability
    on Probability_snp_phe_pop.Probability_snp_phe_pop_id =
        Reference_Probability.Probability_snp_phe_pop_id
    where Phenotype_id = @phe and SNP_id = @s and Population_ID = @pop and
        Bib_ref_id = @bib)
    --Si la cantidad es mayor a 1, Se imprime un mensaje dando una notificacion y con
    --un rollback se cancela la insercion o actualizacion que se esta ejecutando.
    if @cont > 1
    begin
        print 'Esta referencia ya tiene registrada una probabilidad con estas mismas
            características.'
        rollback
    end
end
```

¹⁷ Es una tabla que se crea temporalmente y guarda el registro que se acaba de insertar.

- Las variaciones de tipo SNP y las variaciones de tipo CVN deben estar definidas como polimorfismo.
- Las variaciones no pueden ser SNP y CVN al mismo tiempo.

Para cumplir con estas restricciones hemos creado varios triggers en distintas tablas.

Este primer trigger está especificado en la tabla **SNP**, y se ejecutará al momento de hacer un insert o un update.

```
CREATE TRIGGER [dbo].[tr_Polymorphism_SNP_Only]
ON [dbo].[SNP]
AFTER INSERT, UPDATE
AS
begin
    --Para guardar en la tabla (SNP) solo las variaciones que son de tipo polimorfismo.
    --Si una variación tiene "false" en Frequency_variation_type es porque no es un
    Polimorfismo.
    if (select Frequency_variation_type from Variation where Variation_id = (select
    Variation_id from inserted)) = 'false'
    begin
        print 'Solo se admiten variaciones de tipo polimorfismo.'
        rollback
    end

    --Cuando se hace un insert o un update en la tabla SNP se verifica que las
    --variaciones no estén como SNP y CVN al mismo tiempo.

    --Se toma la cantidad de registros que hay en la tabla CVN cuyo id de variación sea
    --igual al id de variación que estamos insertando o actualizando en la tabla SNP.
    --Si el valor es mayor a cero es porque en la tabla CVN ya está insertado, de manera
    --que imprimimos un mensaje para notificarlo y hacemos un rollback para cancelar
    --el insert o el update que se está ejecutando.
    if (select count(*) from CVN where Variation_id = (select
    Variation_id from inserted)) > 0
    begin
        print 'Esta variación está marcada como CVN. No es posible realizar la
        acción.'
        rollback
    end
end
```

La siguiente restricción está especificada en la tabla **CVN**, y se ejecutará al momento de hacer un insert o un update.

```
CREATE TRIGGER [dbo].[tr_Polymorphism_CVN_Only]
ON [dbo].[CVN]
AFTER INSERT, UPDATE
AS
begin
    --Para guardar solo las variaciones que son de tipo polimorfismo

    --Si una variacion tiene "false" en Frequency_variation_type es porque no es un
    polimorfismo
    if (select Frequency_variation_type from Variation where Variation_id = (select
    Variation_id from inserted)) = 'false'
    begin
        print 'Solo se admiten variaciones de tipo polimorfismo.'
        rollback
    end

    --Cuando se hace un insert o un update en la tabla CVN se verifica que las
    --variaciones no esten como SNP y CVN al mismo tiempo.

    --Se toma la cantidad de registros que hay en la tabla SNP cuyo id de variacion sea
    --igual al id de variacion que estamos insertando o actualizando en la tabla CVN.
    --Si el valor es mayor a cero es porque en la tabla SNP ya esta insertado, de manera
    --que imprimimos un mensajes para notificarlo y hacemos un rollback para cancelar
    --el insert o el update que se esta ejecutando.
    if (select count(*) from SNP where Variation_id = (select Variation_id
    from inserted)) > 0
    begin
        print 'Esta Variacion esta marcada como SNP. No es posible realizar la
        accion.'
        rollback
    end
end
```


La siguiente restricción está especificada en la tabla **Variation**, y se ejecutará al momento de hacer un update.

```
CREATE TRIGGER [dbo].[tr_Frequency_variation_type_change]
ON [dbo].[Variation]
AFTER UPDATE
AS
begin
    --Cuando una variacion esta especificada como (polimorfismo) y está definida en un
    SNP o CNV, no se puede cambiar su tipo a (Mutaicon)

    --Si se actualiza el campo Frequency_variation_type a false (indica que la variacion
    es una mutacion) y en la tabla SNP existe algun registro con variation_id que
    estamos actualizando, se imprime un mensaje notificando la situacion y se hace un
    rollback
    if (Select count(*) from SNP where Variation_id = (select variation_id from
    inserted)) > 0 and (select Frequency_variation_type from inserted) = 'false'
    begin
        print 'Esta variacion esta siendo utilizada como SNP. No es posible cambiar
        el tipo de refernecia de la variacion.'
        rollback
    end

    --Si se actualiza el campo Frequency_variation_type a false (indica que la variacion
    es una mutacion) y en la tabla CVN existe algun registro con variation_id que
    estamos actualizando, se imprime un mensaje notificando la situacion y se hace un
    rollback
    if (Select count(*) from CVN where Variation_id = (select variation_id from
    inserted)) > 0 and (select Frequency_variation_type from inserted) = 'false'
    begin
        print 'Esta variacion esta siendo utilizada como CVN. No es posible cambiar
        el tipo de refernecia de la variacion.'
        rollback
    end
end
```

- Los valores de los casos y los controles en una probabilidad deben ser mayor a cero.

La siguiente restriccion se especifica en la tabla **Probability_snp_phe_pop** y se ejecuta despues de un insert o un update

```
CREATE TRIGGER [dbo].[tr_Case_Control_Validation]
ON [dbo].[Probability_snp_phe_pop]
AFTER INSERT, UPDATE
AS
begin
    --Los datos de los casos y los controles no se pueden establecer menor o igual a cero
    --Contamos los registros que contienen valores mnores o iguales a cero en los datos
    --de los casos y los controles
    --Si se encuentra algun registro, se cancela la transaccion
    if (select count(*) from inserted
        where Case_risk_allele = 0 or Case_reference_allele = 0 or
        Control_risk_allele = 0 or Control_reference_allele = 0) > 0
    begin
        print 'Los valores de los casos y los controles no pueden ser menor o igual a
            cero.'
        rollback
    end
end
```

Procedimientos almacenados

A continuación crearemos los procedimientos almacenados que nos permitirán obtener el OR y el IC, normalmente estos valores vienen especificados en los papers y no sería necesario calcularlos, pero en caso de que el paper no especifique esos valores podemos calcularlos con los procedimientos almacenados a través de los datos utilizados en el estudio comentado en el paper, esto es: El total de personas enfermas con el alelo de riesgo, el total de enfermos con el alelo de referencia, el total de sanos con el alelo de riesgo y el total de sanos con el alelo de referencia.

Anteriormente especificamos los procesos para calcular estos valores manualmente, así que hemos de especificar los mismos pasos en los nuevos procedimientos almacenados.

Estos procedimientos almacenados no se ejecutaran automáticamente como lo hace un trigger, sino que tendrán que ser invocados en el momento que se necesiten. Al invocar estos procedimientos se les indicará si se desea que actúen sobre un registro de la tabla `probability_snp_phe_pop` en concreto o sobre todos los registros de dicha tabla que pertenecen a un SNP determinado.

Calculo de OR en la base de datos a través de un procedimiento almacenado

```
CREATE Procedure [dbo].[Pr_Odds_Ratio]
```

```
@ProbaSNP varchar(20), @Id int
```

```
as
```

```
--Para validar que en la variable @ProbaSNP se introduzca solo "probability" o "snp" esto es para determinar si el calculo se le hara a  
--una probabilidad en concreto o al conjunto de probabilidades de un SNP determinado
```

```
if @ProbaSNP <> 'probability' and @ProbaSNP <> 'snp'
```

```
begin
```

```
    print 'No se ha especificado si es una probabilidad o a un snp. Para una probabilidad se especifica "probability" y para un snp se  
        especifica "snp"'
```

```
    rollback
```

```
end
```

```
--Cuado se aplica a una probabilidad
```

```
if @ProbaSNP = 'probability'
```

```
begin
```

```
--Validar que existen registros con el id especificad haciendo un conteo de los registros.
```

```
if (select count(*) from Probability_snp_phe_pop where Probability_snp_phe_pop_id = @Id) = 0
```

```
begin
```

```
    print 'No existe ninguna probabilidad con el Id indicado'
```

```
    rollback
```

```
end
```

```
--calculamos el odds ratio de la probabilidad indicada aplicando la formula a los datos de case y control.
```

```
Update Probability_snp_phe_pop set Odds_Ratio = ((cast18(Case_risk_allele as numeric(18, 2)) / cast(Case_reference_allele as  
    numeric(18,2))) / (cast (Control_risk_allele as numeric(18, 2)) / cast(Control_reference_allele as numeric(18,2))))
```

```
Where Probability_snp_phe_pop_id = @Id
```

```
end
```

¹⁸ Se utiliza para convertir de un tipo de datos a otro.

```

--Cuando se aplica a todos los registros de probabilidad de un SNP
if @ProbaSNP = 'snp'
begin
    --Validamos que el SNP indicado contenga al menos un registro de probabilidad.
    if (select count(*) from Probability_snp_phe_pop where SNP_id = @Id) = 0
    begin
        print 'El snp indicado no contiene ningun registro de probabilidad'
        rollback
    end
    --Se calcula el Odds Ratio de cada registro de probabilidad que tiene en
    --SNP indicado utilizando los datos de los case y control.
    Update Probability_snp_phe_pop set Odds_Ratio = ((cast(Case_risk_allele as numeric(18,2)) / cast(Case_reference_allele as
numeric(18,2))) / (cast(Control_risk_allele as numeric(18,2)) / cast(Control_reference_allele as numeric(18,2))))
    Where SNP_id = @Id

end

```

Para ejecutar el procedimiento almacena anterior se debe hacer de la siguiente manera:

```
execute Pr_Odds_Ratio 'probability','snp', Id
```

Execute: Comando para ejecutar el procedimiento almacenado.

Pr_Odds_Ratio: Nombre del procedimiento almacenado que se quiere ejecutar.

probability: Especifica que se aplicará a un registro de la tabla probability_snp_phe_pop.

snp: Especifica que se aplicará a todos los registros de la tabla probability_snp_phe_pop que pertenecen a un SNP.

Id: Aquí se especifica el id de la probabilidad o el id del SNP según corresponda.

Ejemplo:

1. `execute Pr_Odds_Ratio 'probability', 15`
Se aplicará solo a la probabilidad 15.
2. `execute Pr_Odds_Ratio 'snp', 4`
Se aplicara a todas las probabilidades del SNP 4

Calculo del Intervalo de Confianza en la base de datos a través de un procedimiento almacenado

```
CREATE Procedure [dbo].[Pr_Intervalo_Confianza]
```

```
@ProbaSNP varchar(20), @Id int
```

```
as
```

```
--Para validar que en la variable @ProbaSNP se introduzca solo "probability" o "snp"
```

```
--Esto es para determinar si el calculo se le hara a una probabilidad en concreto o
```

```
--al conjunto de probabilidades de un SNP determinado
```

```
if @ProbaSNP <> 'probability' and @ProbaSNP <> 'snp'
```

```
begin
```

```
    print 'No se ha especificado si es una probabilidad o a un snp. Para una probabilidad se especifica "probability" y para  
          un snp se especifica "snp"'
```

```
    rollback
```

```
end
```

```
--Declaracion de las variables que se utilizan para calcular el IC.
```

```
--Estas variables van de acuerdo a las formulas.
```

```
declare @a numeric(18,2)
```

```
declare @b numeric(18,2)
```

```
declare @c numeric(18,2)
```

```
declare @d numeric(18,2)
```

```
declare @n0 numeric(18,2)
```

```
declare @n1 numeric(18,2)
```

```
declare @m0 numeric(18,2)
```

```
declare @m1 numeric(18,2)
```

```
declare @t numeric(18,2)
```

```
declare @or numeric(18,2)
```

```
declare @x numeric(18,2)
```

```
declare @IC_Sup numeric(18,2)
```

```
declare @IC_Inf numeric(18,2)
```

--Cuado se aplica a una probabilidad en concreto

if @ProbaSNP = 'probability'

begin

--Validar que existen registros con el id especificado haciendo un conteo de los registros

if (select count(*) from Probability_snp_phe_pop where Probability_snp_phe_pop_id = @Id) = 0

begin

print 'No existe ninguna probabilidad con el Id indicado'

rollback

end

--calculamos el intervalo de confianza aplicando la formula paso a paso.

--Asignar a las variables los valores de los casos y controles

select @or = Odds_Ratio, @a = Case_risk_allele, @b = Control_risk_allele, @c = Case_reference_allele, @d =

Control_reference_allele from Probability_snp_phe_pop

where Probability_snp_phe_pop_id = @Id

set @n1 = @a + @c

set @n0 = @b + @d

set @m1 = @a + @b

set @m0 = @c + @d

set @t = @n1 + @n0

--Hacer el calculo de X

set @x = ((@t - 1)* power¹⁹((@a * @d - @b * @c),2))/(@n1 * @n0 * @m1 * @m0)

set @x = SQRT²⁰(@x)

--Calcular los limites del IC

set @IC_Sup = power(@or,1+(1.96/@x))

set @IC_Inf = power(@or,1-(1.96/@x))

¹⁹ Es una función que devuelve la potencia deseada de un número. Eje.: $5^2 \Rightarrow \text{Power}(5,2)$.

²⁰ Esta función devuelve la raíz cuadrada de un número. Eje.: $\sqrt{25} \Rightarrow \text{SQRT}(25)$

```

--Actualizar la base de datos
update Probability_snp_phe_pop set CI = '95% (' + cast(@IC_Inf as varchar(10))+ ' - ' + cast(@IC_Sup as
                                                                    varchar(10))+ ')'
where Probability_snp_phe_pop_id = @Id

end

--Cuando se aplica a todos los registros de probabilidad de un SNP determinado
if @ProbaSNP = 'snp'
begin
    --Validamos que el SNP indicado contenga al menos un registro de probabilidad.
    if (select count(*) from Probability_snp_phe_pop where SNP_id = @Id) = 0
    begin
        print 'El snp indicado no contiene ningun registro de probabilidad'
        rollback
    end

    --Creamos un bucle para ir calculando el intervalo a cada registro de Probability_snp_phe_pop
    --que pertenece al SNP indicado

    --Para esto primero creamos una variable de tipo tabla para obtener todos los IDs de las probabilidades
    --del SNP indicado
    declare @tabla table (Id int IDENTITY(1,1) , Id_Probability int)

    declare @cont1 int --contador del bucle.
    declare @cont2 int --Limite del bucle.
    declare @Id_Prob int --Para el Probability_snp_phe_pop_id que se este tratando en cada momento.

    --Insertar los datos en la variable "tabla" a partir de la tabla Probability_snp_phe_pop
    insert into @tabla (Id_Probability)
    select Probability_snp_phe_pop_id from Probability_snp_phe_pop
    where SNP_id = @Id

```


--Se inicializan las variables

```
set @cont1 = 0
```

```
set @cont2 = (select count(*) from @tabla) --para la cantidad de registro del SNP
```

--Se crea el bucle

```
while @cont1 < @cont2
```

```
begin
```

```
    set @cont1= @cont1 + 1
```

--Se obtiene el id de una probabilidad en cada iteracion del bucle

```
    set @Id_Prob = (select Id_Probability from @tabla where id = @cont1)
```

--calculamos el intervalo de confianza aplicando la formula paso a paso.

--Asignar los valores de los casos y controles

```
select @or = Odds_Ratio, @a = Case_risk_allele, @b = Control_risk_allele, @c = Case_reference_allele, @d =  
        Control_reference_allele from Probability_snp_phe_pop
```

```
where Probability_snp_phe_pop_id = @Id_Prob
```

```
set @n1 = @a + @c
```

```
set @n0 = @b + @d
```

```
set @m1 = @a + @b
```

```
set @m0 = @c + @d
```

```
set @t = @n1 + @n0
```

--Hacer el calculo de X

```
set @x = ((@t - 1)* power((@a * @d - @b * @c),2))/(@n1 * @n0 * @m1 * @m0)
```

```
set @x = SQRT(@x)
```

--Calcular los limites del IC

```
set @IC_Sup = power(@or,1+(1.96/@x))
```

```
set @IC_Inf = power(@or,1-(1.96/@x))
```

```

--Actualizar la base de datos
update Probability_snp_phe_pop set CI = '95% (' + cast(@IC_Inf as varchar(10)) + ' - ' + cast(@IC_Sup as
                                                                    varchar(10)) + ')'
where Probability_snp_phe_pop_id = @Id_Prob
--El bucle se repita hasta recorrer todos los registros de probabilidad del SNP indicado.
end
end

```

Este procedimiento almacenado se ejecuta de la misma forma que el anterior.

Ejemplo:

1. `execute Pr_Intervalo_Confianza 'probability', 15`
Se aplicará solo a la probabilidad 15.
2. `execute Pr_Intervalo_Confianza 'snp', 4`
Se aplicara a todas las probabilidades del SNP 4

4.5 Explotación de contenidos

Los contenidos serán explotados en una plataforma que ofrece test genéticos directos al consumidor llamada **geneslove.me**. Esta, fue creada por el Centro de Investigación en Métodos de Producción de Software (Pros) y la empresa Gembiosoft. Tiene como objetivo hacer de la información genética una herramienta accesible y eficaz.

Geneslove.me tiene como misión el ofrecer la posibilidad de conocer el perfil genético para diferentes características físicas o enfermedades a través de análisis genéticos contrastados científicamente y analizados con las últimas tecnologías informáticas.



Geneslove.me ofrece una serie de test genéticos entre los cuales está la Alopecia Androgénica, este servicio se ha desarrollado a partir de los datos, del modelo conceptual y del esquema de base de datos que hemos estudiado y desarrollado en esta tesis.

El proceso a seguir para la solicitud del test genético de la Alopecia Androgénica es el siguiente.

En la página principal de geneslove.me nos dirigimos a la Alopecia Androgénica y hacemos click en Comprar.



Se puede especificar la cantidad de test que se desea, introducir un cupón de descuento y añadir otro test. En esta misma página te solicita registrarse, en caso de que no se esté registrado, para pasar al proceso de pago y envío de kit.

PRODUCTO	DESCRIPCIÓN	PRECIO UNITARIO	CANT.	TOTAL	QUITAR TEST
	Alopecia Androgénica	400,00 €	<input type="text" value="1"/>  	400,00 €	

Envío
Subtotal:
IVA 21%:

Envío gratuito!
330,58 €
69,42 €

CUPÓN DESCUENTO

CANJEAR

TOTAL: 400,00 €

AÑADIR OTRO TEST

5. Conclusión

En esta tesis hemos modelado e implementado las probabilidades de riesgo permitiendo que el modelo conceptual del genoma humano sea más completo y adaptado a los problemas actuales. Además modificamos la base de datos y presentamos un esquema que permite la manipulación de los datos probabilísticos de las variaciones de forma sencilla.

Poder determinar el nivel de predisposición a padecer una enfermedad genética, en nuestro caso la Alopecia Androgenética, es muy importante a la hora de tomar decisiones sobre los tratamientos médicos adecuados para prevenir o contrarrestar los efectos de la enfermedad.

Una medida muy utilizada para representar un riesgo es el Odds Ratio (OR), a través de este hemos desarrollado el cálculo de las probabilidades de riesgo. En la actualidad los cálculos estadísticos para determinar las probabilidades de sufrir una enfermedad genética, no están totalmente claros ya que existen algunas discrepancias al respecto, pero creemos que en un futuro no lejano estos cálculos estarán más refinados y mejor establecidos.

Aunque hemos utilizado la Alopecia Androgenética en el desarrollo de esta tesis, queremos enfatizar que todo el proceso para determinar y especificar las probabilidades de riesgo es aplicable a otras enfermedades genéticas.

Con el tiempo las investigaciones en el campo del genoma humano avanza a pasos agigantados, los estudios en los genes y sus variaciones son cada vez más comunes y más fiables, por esta razón esta tesis deja el tema de las probabilidades de riesgo abierto a futuras mejoras que nos permitan estar a la vanguardia ofreciendo información fiable sobre el riesgo de sufrir una enfermedad escrita en los genes.

Trabajos futuros

Como trabajo futuro se necesita determinar la probabilidad que tiene una persona para padecer una enfermedad genética si en sus genes existen varias variaciones asociadas a dicha enfermedad.

Esto se lograría de dos formas

1. Utilizando modelos estadísticos que determinarían el peso que brinda cada variación para la manifestación de la enfermedad.
2. Estudiando subconjuntos de SNPs relacionado con una enfermedad genética, de manera que podríamos obtener un Odds Ratio (OR) para cada subconjunto.
 - Se obtendría una nueva clase **SubSet** que tendrá una relación de mucho a mucho con la clase **SNP** y una relación de uno a mucho con la clase **Probability_snp_phe_pop**.

6. Referencias bibliográficas

1. **Axel M. Hillmer; Felix F. Brockschmidt¹; Sandra Hanneken; Sibylle Eigelshoven; Michael Steffens; Antonia Flaquer; Stefan Herms¹; Tim Becker; Anne-Katrin Kortüm; Dale R. Nyholt; Zhen Zhen Zhao; Grant W. Montgomery; Nicholas G. Martin; Thomas W. Mühleisen¹; Margrieta A. Alblas¹; Susanne Moebus; Karl-Heinz Jöckel; Martina Bröcker-Preuss; Raimund Erbel; Roman Reinartz¹; Regina C. Betz; Sven Cichon¹; Peter Propping; Max P. Baur; Thomas F. Wienker; Roland Kruse; Markus M. Nöthen¹.** A genome-wide association scan identifies new susceptibility variants for male pattern baldness on chromosome 20p11. (2008).
2. **Amanda M. Casto; Brenna M. Henn; Jeffery M. Kidd; Carlos D. Bustamante; Marcus W. Feldman.** A Tale of Two Haplotypes: The *EDA2R/AR* Intergenic Region is the most Divergent Genomic Segment between Africans and East Asians in the Human Genome. (2012)
3. **Justine A. Ellis; Katrina J. Scurrah; Joanna E. Cobb; Sophie G. Zaloumis; Anna E. Duncan; Stephen B. Harrap.** Baldness and the androgen receptor: the AR polyglycine repeat polymorphism does not confer susceptibility to androgenética alopecia. (18 August 2006 / Accepted: 18 December 2006 / Published online: 26 January 2007 Springer-Verlag 2007).
4. **Argenziano, G., Longo, C., & Cameron, A. (2011).** Editor's Choice December 2011 T. Bleiker.
5. **Prodi, D. A., Pirastu, N., Maninchedda, G., Sassu, A., Picciau, A., Palmas, M. A.,... & Pirastu, M. (2008).** EDA2R is associated with androgenetic alopecia. *Journal of investigative dermatology*, 128(9), 2268-2270.
6. **Cobb, J. E., Zaloumis, S. G., Scurrah, K. J., Harrap, S. B., & Ellis, J. A. (2010).** Evidence for two independent functional variants for androgenetic alopecia around the androgen receptor gene. *Experimental dermatology*, 19(11), 1026-1028.
7. **Hillmer, A. M., Hanneken, S., Ritzmann, S., Becker, T., Freudenberg, J., Brockschmidt, F. F., ... & Nöthen, M. M. (2005).** Genetic variation in the human androgen receptor gene is the major determinant of common early-onset androgenetic alopecia. *The American Journal of Human Genetics*, 77(1), 140-148.
8. **Richards, J. B., Yuan, X., Geller, F., Waterworth, D., Bataille, V., Glass, D., ... & Mooser, V. (2008).** Male-pattern baldness susceptibility locus at 20p11. *Nature genetics*, 40(11), 1282-1284.

9. **Li, R., Brockschmidt, F. F., Kiefer, A. K., Stefansson, H., Nyholt, D. R., Song, K., ... & Richards, J. B. (2012).** Six novel susceptibility Loci for early-onset androgenetic alopecia and their unexpected association with common diseases. *PLoS genetics*, 8(5), e1002746.
10. **Hillmer, A. M., Brockschmidt, F. F., Hanneken, S., Eigelshoven, S., Steffens, M., Flaquer, A., ... & Nöthen, M. M. (2008).** Susceptibility variants for male-pattern baldness on chromosome 20p11. *Nature genetics*, 40(11), 1279-1281.
11. **Brockschmidt, F. F., Heilmann, S., Ellis, J. A., Eigelshoven, S., Hanneken, S., Herold, C., ... & Nöthen, M. M. (2011).** Susceptibility variants on chromosome 7p21. 1 suggest HDAC9 as a new candidate gene for male-pattern baldness. *British Journal of Dermatology*, 165(6), 1293-1302.
12. **Brockschmidt, F. F., Hillmer, A. M., Eigelshoven, S., Hanneken, S., Heilmann, S., Barth, S., ... & Nöthen, M. M. (2010).** Fine mapping of the human AR/EDA2R locus in androgenetic alopecia. *British Journal of Dermatology*, 162(4), 899-903.
13. **Hillmer, A. M., Freudenberg, J., Myles, S., Herms, S., Tang, K., Hughes, D. A., ... & Nöthen, M. M. (2009).** Recent positive selection of a human androgen receptor/ectodysplasin A2 receptor haplotype and its relationship to male pattern baldness. *Human genetics*, 126(2), 255-264.
14. **Flórez, J., & Troncoso, M. V. (1991).** Síndrome de Down. Avances en acción familiar. Fundación síndrome de Down de Cantabria. Santander.
15. **Kaufman, KD, Olsen, EA, Whiting, D., Savin, R., Devillez, R., Bergfeld, W., Y Gormley, GJ (1998).** La finasterida en el tratamiento de hombres con alopecia androgenética. *Revista de la Academia Americana de Dermatología*, 39 (4), 578-589.
16. **Struewing, J. P., Hartge, P., Wacholder, S., Baker, S. M., Berlin, M., McAdams, M., ... & Tucker, M. A. (1997).** The risk of cancer associated with specific mutations of BRCA1 and BRCA2 among Ashkenazi Jews. *New England Journal of Medicine*, 336(20), 1401-1408
17. **Yanguas, J. J. (2007).** Modelo de atención a las personas con enfermedad de Alzheimer (No. 21011). Imsero.
18. **Peña-Casanova, J. (1999).** Enfermedad de Alzheimer. Sociedad Española de Neurología. Guías en Demencias. Fundación «La Caixa», Barcelona.

- 19. Capone, G., Goyal, P., Ares, W., & Lannigan, E. (2006, August).** Neurobehavioral disorders in children, adolescents, and young adults with Down syndrome. In *American Journal of Medical Genetics Part C: Seminars in Medical Genetics* (Vol. 142, No. 3, pp. 158-172). Wiley Subscription Services, Inc., A Wiley Company.
- 20. Flórez, J., & Troncoso, M. V. (1991).** Síndrome de Down. Avances en acción familiar. Fundación síndrome de Down de Cantabria. Santander.
- 21. Edenberg, H. J., Dick, D. M., Xuei, X., Tian, H., Almasy, L., Bauer, L. O., ... & Begleiter, H. (2004).** Variations in GABRA2, Encoding the $\alpha 2$ Subunit of the GABA_A Receptor, Are Associated with Alcohol Dependence and with Brain Oscillations. *The American Journal of Human Genetics*, 74(4), 705-714.

7. Anexos



QUIÉNES SOMOS
QUÉ OFRECEMOS
CATÁLOGO DE PRODUCTOS






ALOPECIA ANDROGÉNICA

Con éste Test Genético te ayudamos a que sepas si tienes predisposición a padecer Alopecia Androgénica.

Más información

La Alopecia Androgénica, conocida como calvicie común; consiste en la caída del pelo de manera progresiva, llegando en algunos casos a dejar el cuero cabelludo al descubierto.

El test ofrecido por **geneslove.me** para la detección de la alopecia androgénica consiste en que realices una prueba "muy sencilla" de tu ADN. A través de ésta prueba se te ayudará a determinar si has heredado o no, el gen receptor de la alopecia androgénica. Los resultados de ésta predisposición se te harán llegar de forma totalmente segura y privada sólo a ti. La relevancia de ésta prueba es que puedes conocer tu predisposición a padecerla aunque hasta éste momento no exista ningún síntoma de pérdida de cabello.

La importancia de conocer si genéticamente tienes alopecia androgénica; ante todo es estar bien informado cuanto antes. Y no perder el tiempo, haciendo uso de tratamientos que no tengan gran eficacia.

El test genético no tiene ningún riesgo, es decir, es completamente seguro. El resultado es realizado por laboratorios genéticos de gran prestigio internacional, quienes garantizan en todo momento la calidad del resultado así como el cumplimiento de la legislación en materia de protección de datos.

Variaciones de la AGA

Genetic Variant	Chr.	Position*	EA ^b /NEA	EA ^f	OR (95% CI)	p value	Gene ^c	Number of GW Significant SNPs	\hat{r}
rs12565727	1	10955669	A/G	0.789	1.33 (1.22–1.45)	9.07×10^{-11}	TARDBP	8	<0.01
rs9287638	2	239359379	A/C	0.562	1.31 (1.21–1.41)	1.01×10^{-12}	HDAC4	9	0.10
rs2073963	7	18844399	G/T	0.530	1.29 (1.20–1.38)	1.08×10^{-12}	HDAC9	30	<0.01
rs6945541	7	68249896	C/T	0.539	1.27 (1.18–1.38)	1.71×10^{-9}	AUTS2	6	<0.01
rs12373124	17	41279999	T/C	0.438	1.33 (1.21–1.45)	5.07×10^{-10}	17q21.31	118	0.24
rs10502861	18	41054146	C/T	0.775	1.28 (1.18–1.39)	2.62×10^{-9}	SETBP1	16	0.18
rs6047844	20	21985575	T/C	0.460	1.60 (1.49–1.72)	1.71×10^{-39}	PAX1, FOXA2	277	<0.01
rs2497938	X	66479743	T/C	0.850	2.20 (2.04–2.37)	2.40×10^{-91}	AR	181	0.32

65

Síndrome de Down



Alzheimer

Cerebro normal



Cerebro con Alzheimer

